(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 October 2001 (18.10.2001)

PCT

(10) International Publication Number
WO 01/77311 A2

(51) International Patent Classification[7]: C12N 15/00

(21) International Application Number: PCT/US01/11393

(22) International Filing Date: 6 April 2001 (06.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/196,001 7 April 2000 (07.04.2000) US

(71) Applicant *(for all designated States except US)*: **BASF PLANT SCIENCE GMBH** [DE/DE]; D-67056 Ludwigshafen (DE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **DA COSTA SILVA, Oswaldo** [BR/US]; 203 Littleford Lane, Apex, NC 27502 (US). **BOHNERT, Hans, J.** [US/US]; 26 Davis Drive, Research Triangle Park, NC 27709 (US). **VAN THIELEN, Nocha** [CA/US]; 206-T Hyde Park Court, Cary, NC 27513 (US). **CHEN, Ruoying** [CN/US]; 105 Collier Place - 1G, Cary, NC 27513 (US).

(74) Agents: **COBERN, Lisa, M.** et al.; Sutherland Asbill & Brennan, LLP, 999 Peachtree Street, N.E., Atlanta, GA 30309-3996 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European

[Continued on next page]

(54) Title: TRANSCRIPTION FACTOR STRESS-RELATED PROTEINS AND METHODS OF USE IN PLANTS

(57) Abstract: A transgenic plant transformed by a Transcription Factor Stress-Related Protein (TFSRP) coding nucleic acid, wherein expression of the nucleic acid sequence in the plant results in increased tolerance to environmental stress as compared to a wild type variety of the plant. Also provided are agricultural products, including seeds, produced by the transgenic plants. Also provided are isolated TFSRPs, and isolated nucleic acid coding TFSRPs, and vectors and host cells containing the latter.

Nucleotide sequence of the partial APS-2 from *Physcomitrella patens* (SEQ ID NO:1)

TCAAGCCACTCATCCGAGCATAGAACATCACAACCCACCTTGATGATCATTCTCTCA
GCCGACCAGCGTCAATTACGCTGCGTATCGCTCTAGCTTGAGGAAGGCACCCTCGCC
CTCTTCGCCGCGGAAGTAGCCCTCTGCTTCACGAGGGCGGCAAAACTCTCCCAAGGC
AGTTCCGGGGGGATGGGATATAGCTGCAGCTGCTGTGGGGAATCCTCAAAATTGTA
CGGGATCTTCTTCTTTGTGTAGAAGATGCCAACATCGTAGGCCCGGGCAGCTTCTTC
CGGAGTTTCATATGTTCCCAGCCATATCTTACGTTTCTGAGATGTGGGTCGAATTTCT
GTCACCCATTTGTTTAGCTCGGGCCGGTGCCGAACCCCCCTAAAACTGGTCGTATCG
CCAGTGTTGCTAGCAGAAACTTCTTCGGTATCCCATGCCGATGGGGCCTTATTTAAA
TCAATATTCCGAAATTTAAAGGCATTCCGACCGCTAGTGTCTTTCGCCGCTAACCGC
ATTCCCTTGGATTCTTCTTCCAAACTAGATTCAGACTTGCTCTCCTGCCAACTTCTTTT
TTCACTTTCGGGGATTCTATTTTAGTCGTTAACTGCAACGCCTGTTCTTTGACCTTGC
CACCACAAGGATCCCACTTCTTTGTTTTGGGCTTCCCCTGTTCAATAATGCTGGAAAT
TGTCAAATTCATGAACTACCCAATTGCAACCCCTCCCACCGGGATGGATTGATCGCC
AAAATTTCGTAGTAACTTAACTTTCATACAACAACTTGAGTTCCTTCGCTATTAGGG
ACACGTGGCAGAAACTTGGACGTGCAAGCGTATGTACTCATCAGAGTTTGACAGCG
CATAAAATCATATAAAAAGTCTTGAAGAAGCGTTGTTTAATTCATGGGTAACCACGA
GTTACGCGGAGCGTCGGCAGCAAGGAGAGGACGACCAGGCGGCAAGAAGATGCGT
CGGCAAGAGCTCGTGC

WO 01/77311 A2

patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TRANSCRIPTION FACTOR STRESS-RELATED PROTEINS AND METHODS OF USE IN PLANTS

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the priority benefit of U.S. Provisional Application Serial No. 60/196,001 filed April 7, 2000.

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] This invention relates generally to nucleic acid sequences encoding proteins that are associated with abiotic stress responses and abiotic stress tolerance in plants. In particular, this invention relates to nucleic acid sequences encoding proteins that confer drought, cold, and/or salt tolerance to plants.

Background Art

[0002] Abiotic environmental stresses, such as drought stress, salinity stress, heat stress, and cold stress, are major limiting factors of plant growth and productivity. Crop losses and crop yield losses of major crops such as rice, maize (corn) and wheat caused by these stresses represent a significant economic and political factor and contribute to food shortages in many underdeveloped countries.

[0003] Plants are typically exposed during their life cycle to conditions of reduced environmental water content. Most plants have evolved strategies to protect themselves against these conditions of desiccation. However, if the severity and duration of the drought conditions are too great, the effects on plant development, growth and yield of most crop plants are profound. Furthermore, most of the crop plants are very susceptible to higher salt concentrations in the soil. Continuous exposure to drought and high salt causes major alterations in the plant metabolism. These great changes in metabolism ultimately lead to cell death and consequently yield losses.

[0004] Developing stress-tolerant plants is a strategy that has the potential to solve or mediate at least some of these problems. However, traditional plant breeding strategies to develop new lines of plants that exhibit resistance (tolerance) to these types of stresses are relatively slow and require specific resistant lines for crossing with the desired line. Limited germplasm resources for stress tolerance and incompatibility in crosses between distantly related plant species represent significant problems encountered in conventional breeding. Additionally, the cellular processes leading to drought, cold and salt tolerance in model, drought- and/or salt-tolerant plants are complex in nature and involve multiple mechanisms of cellular adaptation and numerous metabolic pathways. This multi-component nature of stress tolerance has not only made breeding for tolerance largely unsuccessful, but has also limited the ability to genetically engineer stress tolerance plants using biotechnological methods.

[0005] Therefore, what is needed is the identification of the genes and proteins involved in these multi-component processes leading to stress tolerance. Elucidating the function of genes expressed in stress tolerant plants will not only advance our understanding of plant adaptation and tolerance to environmental stresses, but also may provide important information for designing new strategies for crop improvement.

[0006] One model plant used in the study of stress tolerance is *Arabidopsis thaliana*. There are at least four different signal-transduction pathways leading to stress tolerance in the model plant *Arabidopsis thaliana*. These pathways are under the control of distinct transcription factors (Shinozaki et al., 2000 Curr. Op. Pl. Biol. 3:217-23). Regulators of genes, especially transcription factors, involved in these tolerance pathways are particularly suitable for engineering tolerance into plants because a single gene can activate a whole cascade of genes leading to the tolerant phenotype. Consequently, transcription factors are important targets in the quest to identify genes conferring stress tolerance to plants.

[0007] One transcription factor that has been identified in the prior art is the *Arabidopsis thaliana* transcription factor CBF (Jaglo-Ottosen et al., 1998 Science 280:104-6). Over-expression of this gene in *Arabidopsis* conferred drought tolerance to this plant (Kasuga et al., 1999 Nature Biotech. 17:287-91). However, CBF is the only example to date of a transcription factor able to confer drought tolerance to plants upon over-expression.

[0008] Although some genes that are involved in stress responses in plants have been characterized, the characterization and cloning of plant genes that confer stress tolerance remains largely incomplete and fragmented. For example, certain studies have indicated that drought and salt stress in some plants may be due to additive gene effects, in contrast to other

research that indicates specific genes are transcriptionally activated in vegetative tissue of plants under osmotic stress conditions. Although it is generally assumed that stress-induced proteins have a role in tolerance, direct evidence is still lacking, and the functions of many stress-responsive genes are unknown.

[0009] There is a need, therefore, to identify genes expressed in stress tolerant plants that have the capacity to confer stress resistance to its host plant and to other plant species. Newly generated stress tolerant plants will have many advantages, such as increasing the range that crop plants can be cultivated by, for example, decreasing the water requirements of a plant species.

SUMMARY OF THE INVENTION

[0010] This invention fulfills in part the need to identify new, unique transcription factors capable of conferring stress tolerance to plants upon over-expression. The present invention provides a transgenic plant cell transformed by a Transcription Factor Stress-Related Protein (TFSRP) coding nucleic acid, wherein expression of the nucleic acid sequence in the plant cell results in increased tolerance to environmental stress as compared to a wild type variety of the plant cell. Namely, described herein are the transcription factors 1) CAAT-Box like Binding Factor-3 (CABF-3); 2) Zinc Finger-2 (ZF-2) 3) Zinc Finger-3 (ZF-3); 4) Zinc Finger-4 (ZF-4); 5) Zinc Finger-5 (ZF-5); 6) AP2 Similar Factor-2 (APS-2); 7) Sigma Factor Like Factor-1 (SFL-1); and 8) MYB Factor-1 (MYB-1), all from *Physcomitrella patens*.

[0011] The invention provides in some embodiments that the TFSRP and coding nucleic acid are that found in members of the genus *Physcomitrella*. In another preferred embodiment, the nucleic acid and protein are from a *Physcomitrella patens*. The invention provides that the environmental stress can be salinity, drought, temperature, metal, chemical, pathogenic and oxidative stresses, or combinations thereof. In preferred embodiments, the environmental stress can be drought or cold temperature.

[0012] The invention further provides a seed produced by a transgenic plant transformed by a TFSRP coding nucleic acid, wherein the plant is true breeding for increased tolerance to environmental stress as compared to a wild type variety of the plant. The invention further provides a seed produced by a transgenic plant expressing a TFSRP, wherein the plant is true breeding for increased tolerance to environmental stress as compared to a wild type variety of the plant.

[0013] The invention further provides an agricultural product produced by any of the below-described transgenic plants, plant parts or seeds. The invention further provides an isolated TFSRP as described below. The invention further provides an isolated TFSRP coding nucleic acid, wherein the TFSRP coding nucleic acid codes for a TFSRP as described below.

[0014] The invention further provides an isolated recombinant expression vector comprising a TFSRP coding nucleic acid as described below, wherein expression of the vector in a host cell results in increased tolerance to environmental stress as compared to a wild type variety of the host cell. The invention further provides a host cell containing the vector and a plant containing the host cell.

[0015] The invention further provides a method of producing a transgenic plant with a TFSRP coding nucleic acid, wherein expression of the nucleic acid in the plant results in increased tolerance to environmental stress as compared to a wild type variety of the plant comprising: (a) transforming a plant cell with an expression vector comprising a TFSRP coding nucleic acid, and (b) generating from the plant cell a transgenic plant with an increased tolerance to environmental stress as compared to a wild type variety of the plant. In preferred embodiments, the TFSRP and TFSRP coding nucleic acid are as described below.

[0016] The present invention further provides a method of identifying a novel TFSRP, comprising (a) raising a specific antibody response to a TFSRP, or fragment thereof, as described below; (b) screening putative TFSRP material with the antibody, wherein specific binding of the antibody to the material indicates the presence of a potentially novel TFSRP; and (c) identifying from the bound material a novel TFSRP in comparison to known TFSRP. Alternatively, hybridization with nucleic acid probes as described below can be used to identify novel TFSRP nucleic acids.

[0017] The present invention also provides methods of modifying stress tolerance of a plant comprising, modifying the expression of a TFSRP in the plant, wherein the TFSRP is as described below. The invention provides that this method can be performed such that the stress tolerance is either increased or decreased. Preferably, stress tolerance is increased in a plant via increasing expression of a TFSRP.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] Figures 1(A-H) show the partial cDNA sequences of APS-2 (SEQ ID NO:1), ZF-2 (SEQ ID NO:2), ZF-3 (SEQ ID NO:3), ZF-4 (SEQ ID NO:4), ZF-5 (SEQ ID NO:5),

MYB-1 (SEQ ID NO:6), CABF-3 (SEQ ID NO:7) and SFL-1 (SEQ ID NO:8) from *Physcomitrella patens*.

[0019] Figures 2(A-H) show the full-length cDNA sequences of APS-2 (SEQ ID NO:9), ZF-2 (SEQ ID NO:10), ZF-3 (SEQ ID NO:11), ZF-4 (SEQ ID NO:12), ZF-5 (SEQ ID NO:13), MYB-1 (SEQ ID NO:14), CABF-3 (SEQ ID NO:15) and SFL-1 (SEQ ID NO:16) from *Physcomitrella patens*.

[0020] Figures 3(A-H) show the deduced amino acid sequences of APS-2 (SEQ ID NO:17), ZF-2 (SEQ ID NO:18), ZF-3 (SEQ ID NO:19), ZF-4 (SEQ ID NO:20), ZF-5 (SEQ ID NO:21), MYB-1 (SEQ ID NO:22), CABF-3 (SEQ ID NO:23) and SFL-1 (SEQ ID NO:24) from *Physcomitrella patens*.

[0021] Figure 4 shows a diagram of the plant expression vector pBPSSC022 containing the super promoter driving the expression of SEQ ID NOs: 9, 10, 11, 12, 13, 14, 15, and 16 ("Desired Gene"). The components are: NPTII kanamycin resistance gene (Hajdukiewicz et al. 1994 Pl. Mol Biol. 25:989-98), AtAct2-i promoter (An et al. 1996 Plant J. 10:107-21), OCS3 terminator (Weigel et al. 2000 Pl. Physiol. 122: 1003-13).

[0022] Figure 5 shows the results of a drought stress test with over-expressing PpZF-2 transgenic plants and wild-type *Arabidopsis* lines. The transgenic lines display a tolerant phenotype. Individual transformant lines are shown.

[0023] Figure 6 shows the results of a drought stress test with over-expressing PpZF-3 transgenic plants and wild-type *Arabidopsis* lines. The transgenic lines display a tolerant phenotype. . Individual transformant lines are shown.

[0024] Figure 7 shows the results of a drought stress test with over-expressing PpZF-4 transgenic plants and wild-type *Arabidopsis* lines. The transgenic lines display a tolerant phenotype. Individual transformant lines are shown.

[0025] Figure 8 shows the results of a drought stress test with over-expressing PpZF-5 transgenic plants and wild-type *Arabidopsis* lines. The transgenic lines display a tolerant phenotype. Individual transformant lines are shown.

[0026] Figure 9 shows the results of a drought stress test with over-expressing PpCABF-3 transgenic plants and wild-type *Arabidopsis* lines. The transgenic lines display a tolerant phenotype. Individual transformant lines are shown.

[0027] Figure 10 shows the results of a drought stress test with over-expressing PpAPS-2 transgenic plants and wild-type *Arabidopsis* lines. The transgenic lines display a tolerant phenotype. Individual transformant lines are shown.

[0028] Figure 11 shows the results of a drought stress test with over-expressing PpSFL-1 transgenic plants and wild-type *Arabidopsis* lines. The transgenic lines display a tolerant phenotype. Individual transformant lines are shown.

[0029] Figure 12 shows the results of a drought stress test with over-expressing PpMYB-1 transgenic plants and wild-type *Arabidopsis* lines. The transgenic lines display a tolerant phenotype. Individual transformant lines are shown.

[0030] Figure 13 shows the results of a freezing stress test with over-expressing PpCABF-3 transgenic plants and wild-type *Arabidopsis* lines. The transgenic lines display a tolerant phenotype. Individual transformant lines are shown.

[0031] Figure 14 shows the results of a freezing stress test with over-expressing PpZF-2 transgenic plants and wild-type *Arabidopsis* lines. The transgenic lines display a tolerant phenotype. Individual transformant lines are shown.

[0032] Figure 15 shows the results of a freezing stress test with over-expressing PpZF-3 transgenic plants and wild-type *Arabidopsis* lines. The transgenic lines display a tolerant phenotype. Individual transformant lines are shown.

DETAILED DESCRIPTION OF THE INVENTION

[0033] The present invention may be understood more readily by reference to the following detailed description of the preferred embodiments of the invention and the Examples included herein. However, before the present compounds, compositions, and methods are disclosed and described, it is to be understood that this invention is not limited to specific nucleic acids, specific polypeptides, specific cell types, specific host cells, specific conditions, or specific methods, etc., as such may, of course, vary, and the numerous modifications and variations therein will be apparent to those skilled in the art. It is also to be understood that the terminology used herein is for the purpose of describing specific embodiments only and is not intended to be limiting. In particular, the designation of the amino acid sequences as protein "Transcription Factor Stress-Related Proteins" (TFSRPs), in no way limits the functionality of those sequences.

[0034] The present invention provides a transgenic plant cell transformed by a TFSRP coding nucleic acid, wherein expression of the nucleic acid sequence in the plant cell results in increased tolerance to environmental stress as compared to a wild type variety of the plant cell. The invention further provides transgenic plant parts and transgenic plants containing the plant cells described herein. Also provided is a plant seed produced by a transgenic plant transformed by a TFSRP coding nucleic acid, wherein the seed contains the

TFSRP coding nucleic acid, and wherein the plant is true breeding for increased tolerance to environmental stress as compared to a wild type variety of the plant. The invention further provides a seed produced by a transgenic plant expressing a TFSRP, wherein the seed contains the TFSRP, and wherein the plant is true breeding for increased tolerance to environmental stress as compared to a wild type variety of the plant. The invention also provides an agricultural product produced by any of the below-described transgenic plants, plant parts and plant seeds.

[0035] As used herein, the term "variety" refers to a group of plants within a species that share constant characters that separate them from the typical form and from other possible varieties within that species. While possessing at least one distinctive trait, a variety is also characterized by some variation between individuals within the variety, based primarily on the Mendelian segregation of traits among the progeny of succeeding generations. A variety is considered "true breeding" for a particular trait if it is genetically homozygous for that trait to the extent that, when the true-breeding variety is self-pollinated, a significant amount of independent segregation of the trait among the progeny is not observed. In the present invention, the trait arises from the transgenic expression of one or more DNA sequences introduced into a plant variety.

[0036] The present invention describes for the first time that the *Physcomitrella patens* TFSRPs, APS-2, ZF-2, ZF-3, ZF-4, ZF-5, MYB-1, CABF-3 and SFL-1, are useful for increasing a plant's tolerance to environmental stress. The PpAPS-2 protein (AP2 Similar) contains a region of similarity with the AP2 domain present in some plant transcription factors. Apetala-2 (AP2) is a homeotic gene in *Arabidopsis* and mutations in this gene result in the generation of flowers without petals. The AP2 domain is found in not only homeotic genes in plants, but also in proteins with diverse function.

[0037] Another group of novel predicted proteins described herein are PpZF-2, PpZF-3, PpZF-4 and PpZF-5, which show sequence similarity to the Zinc-Finger class of transcription factors. Zinc-finger transcription factors share in common a specific secondary structure wherein a zinc molecule is sequestered by the interaction with cysteine or histidine amino acid residues. Through these "fingers," the proteins interact with their specific DNA targets and regulate transcription of the target genes. Zinc-finger factors are associated with a multitude of biological phenomena. For example, in yeast zinc fingers are related with the regulation of multiple genes, e.g. genes involved in general metabolism. In plants, a zinc-finger protein, CONSTANS, is responsible for determining flowering time (Putterill et al. 1995 Cell 80:847-57). Sakamoto et al. (2000 Gene 248:23-32) also report the activation of

the gene expression of three zinc finger proteins in *Arabidopsis* during water-stress treatments. They did not, however, present any data linking this increased expression with stress tolerance. Finally, Lippuner et al. (1996 JBC 271:12859-66) have reported that a particular class of zinc-finger proteins was able to confer salt tolerance to yeast mutants, however no data showing increased salt tolerance to whole plants was presented.

[0038] Another novel predicted protein described herein is a PpMYB-1 protein that shares sequence homology with transcription factors from the MYB family. This group of transcription factors have the highest degree of homology in the "MYB domain". In addition to being involved in pigment formation in maize (Shinozaki et al. 2000. Curr. Op. Pl. Biol. 3: 217-23), it has also been proposed that a MYB-containing protein is involved in regulating stress-related gene expression in plants. In particular, a MYB-containing protein, AtMYB2 has been shown to be stress-induced (PCT Application No. WO 99/16878). However, no data has been presented, demonstrating that the over-expression of AtMYB2 leads to stress tolerance in a plant.

[0039] Yet another novel predicted protein described herein is PpCABF-3, which is similar to the domain "B" of other CAAT-Box Binding Factors (Johnson and McKnight. 1989. Ann. Rev. Biochem. 58:799-840). In general, CABFs are parts of multi-component transcription activation complexes and act as general transcriptional regulators and activators. The particular combination of the different CABFs and other sub-units in the complex determines the target genes. PpCABF-3 seems to be important for the activation of stress-related genes upon over-expression in *Arabidopsis thaliana*. PpCABF-3 is homologous to other two CAAT-Box Binding Factors from *Physcomitrella patens*, namely PpCABF-1 and PpCABF-2. Based upon a phylogenic analysis, it is believed that these proteins belong to an exclusive class of CAAT-Box Binding proteins.

[0040] A final group of novel predicted proteins described herein includes the PpSFL-1 (Sigma Factor Like) protein. The SFL-1 shares a high degree of sequence with prokaryotic and plant chloroplast sigma factors. Sigma factors are essential for determining promoter recognition and consequently correct transcription initiation in prokaryotes as well as in chloroplasts. Chloroplasts are a major target for engineering stress tolerance, since these organelles are heavily impaired during stress conditions. Attenuation of chloroplast damage can lead to increased stress tolerance in plants.

[0041] Accordingly, the present invention provides isolated TFSRPs selected from the group consisting of APS-2, ZF-2, ZF-3, ZF-4, ZF-5, MYB-1, CABF-3, SFL-1 and homologs thereof. In preferred embodiments, the TFSRP is selected from 1) a AP2 Similar-2

(APS-2) protein as defined in SEQ ID NO:17; 2) a Zinc-Finger Factor-2 (ZF-2) protein as defined in SEQ ID NO:18; 3) a Zinc-Finger Factor-3 (ZF-3) protein as defined in SEQ ID NO:19; 4) a Zinc-Finger Factor-4 (ZF-4) protein as defined in SEQ ID NO:20; 5) a Zinc-Finger Factor-5 (ZF-5) protein as defined in SEQ ID NO:21; 6) a MYB-1 (MYB-1) protein as defined in SEQ ID NO:22; 7) a CAAT-Box Binding Factor-3 (CABF-3) protein as defined in SEQ ID NO:23; 8) a Sigma Factor Like (SFL-1) protein as defined in SEQ ID NO:24, and homologs and orthologs thereof. Homologs and orthologs of the amino acid sequences are defined below.

[0042] The TFSRPs of the present invention are preferably produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the protein is cloned into an expression vector (as described below), the expression vector is introduced into a host cell (as described below) and the TFSRP is expressed in the host cell. The TFSRP can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Alternative to recombinant expression, a TFSRP polypeptide, or peptide can be synthesized chemically using standard peptide synthesis techniques. Moreover, native TFSRP can be isolated from cells (e.g., *Physcomitrella patens*), for example using an anti-TFSRP antibody, which can be produced by standard techniques utilizing a TFSRP or fragment thereof.

[0043] The invention further provides an isolated TFSRP coding nucleic acid. The present invention includes TFSRP coding nucleic acids that encode TFSRPs as described herein. In preferred embodiments, the TFSRP coding nucleic acid is selected from 1) a AP2 Similar-2 (APS-2) nucleic acid as defined in SEQ ID NO:9; 2) a Zinc-Finger Factor-2 (ZF-2) nucleic acid as defined in SEQ ID NO:10; 3) a Zinc-Finger Factor-3 (ZF-3) nucleic acid as defined in SEQ ID NO:11; 4) a Zinc-Finger Factor-4 (ZF-4) nucleic acid as defined in SEQ ID NO:12; 5) a Zinc-Finger Factor-5 (ZF-5) nucleic acid as defined in SEQ ID NO:13; 6) a MYB-1 nucleic acid as defined in SEQ ID NO:14; 7) a CAAT-Box Binding Factor-3 (CABF-3) nucleic acid as defined in SEQ ID NO:15; 8) a Sigma Factor Like (SFL-1) nucleic acid as defined in SEQ ID NO:16 and homologs and orthologs thereof. Homologs and orthologs of the nucleotide sequences are defined below. In one preferred embodiment, the nucleic acid and protein are isolated from the plant genus *Physcomitrella*. In another preferred embodiment, the nucleic acid and protein are from a *Physcomitrella patens* (*P. patens*) plant.

[0044] As used herein, the term "environmental stress" refers to any sub-optimal growing condition and includes, but is not limited to, sub-optimal conditions associated with salinity, drought, temperature, metal, chemical, pathogenic and oxidative stresses, or

combinations thereof. In preferred embodiments, the environmental stress can be salinity, drought, or temperature, or combinations thereof, and in particular, can be high salinity, low water content or low temperature. It is also to be understood that as used in the specification and in the claims, "a" or "an" can mean one or more, depending upon the context in which it is used. Thus, for example, reference to "a cell" can mean that at least one cell can be utilized.

[0045] As also used herein, the terms "nucleic acid" and "nucleic acid molecule" are intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. This term also encompasses untranslated sequence located at both the 3' and 5' ends of the coding region of the gene: at least about 1000 nucleotides of sequence upstream from the 5' end of the coding region and at least about 200 nucleotides of sequence downstream from the 3' end of the coding region of the gene. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

[0046] An "isolated" nucleic acid molecule is one that is substantially separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of some of the sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated TFSRP nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived (e.g., a *Physcomitrella patens* cell). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be free from some of the other cellular material with which it is naturally associated, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized.

[0047] A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having a nucleotide sequence of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, a *P. patens* TFSRP cDNA can be isolated from a *P. patens* library using all or portion of one of the sequences of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 or SEQ ID NO:8. Moreover, a nucleic acid molecule encompassing all or a portion of one of the sequences of

SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 and SEQ ID NO:8 can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence. For example, mRNA can be isolated from plant cells (e.g., by the guanidinium-thiocyanate extraction procedure of Chirgwin et al., 1979 Biochemistry 18:5294-5299) and cDNA can be prepared using reverse transcriptase (e.g., Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, MD; or AMV reverse transcriptase, available from Seikagaku America, Inc., St. Petersburg, FL). Synthetic oligonucleotide primers for polymerase chain reaction amplification can be designed based upon one of the nucleotide sequences shown in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 and SEQ ID NO:8. A nucleic acid molecule of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid molecule so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to a TFSRP nucleotide sequence can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

[0048] In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises one of the nucleotide sequences shown in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16. These cDNAs comprise sequences encoding the TFSRPs (i.e., the "coding region", indicated in Table 1), as well as 5' untranslated sequences and 3' untranslated sequences. It is to be understood that SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16 comprise both coding regions and 5' and 3' untranslated regions. Alternatively, the nucleic acid molecules of the present invention can comprise only the coding region of any of the sequences in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16 or can contain whole genomic fragments isolated from genomic DNA. A coding region of these sequences is indicated as "ORF position". The present invention also includes TFSRP coding nucleic acids that encode TFSRPs as described herein. Preferred is a TFSRP coding nucleic acid that encodes a TFSRP selected from the group consisting of, APS-2 (SEQ ID NO:17), ZF-2 (SEQ ID NO:18), ZF-3 (SEQ ID NO:19), ZF-4 (SEQ ID NO:20), ZF-5 (SEQ ID NO:21), MYB-1 (SEQ ID NO:22), CABF-3 (SEQ ID NO:23) and SFL-1 (SEQ ID NO:24).

[0049] Moreover, the nucleic acid molecule of the invention can comprise only a portion of the coding region of one of the sequences in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16, for example, a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a TFSRP. The nucleotide sequences determined from the cloning of the TFSRP genes from *P. patens* allow for the generation of probes and primers designed for use in identifying and/or cloning TFSRP homologs in other cell types and organisms, as well as TFSRP homologs from other mosses and related species.

[0050] Portions of proteins encoded by the TFSRP nucleic acid molecules of the invention are preferably biologically active portions of one of the TFSRPs described herein. As used herein, the term "biologically active portion of" a TFSRP is intended to include a portion, e.g., a domain/motif, of a TFSRP that participates in a stress tolerance response in a plant, has an activity as set forth in Table 1, or participates in the transcription of a protein involved in a stress tolerance response in a plant. To determine whether a TFSRP, or a biologically active portion thereof, can participate in transcription of a protein involved in a stress tolerance response in a plant, or whether repression of a TFSRP results in increased stress tolerance in a plant, a stress analysis of a plant comprising the TFSRP may be performed. Such analysis methods are well known to those skilled in the art, as detailed in Example 7. More specifically, nucleic acid fragments encoding biologically active portions of a TFSRP can be prepared by isolating a portion of one of the sequences in SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 and SEQ ID NO:24, expressing the encoded portion of the TFSRP or peptide (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the TFSRP or peptide.

[0051] Biologically active portions of a TFSRP are encompassed by the present invention and include peptides comprising amino acid sequences derived from the amino acid sequence of a TFSRP, e.g., an amino acid sequence of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 or SEQ ID NO:24, or the amino acid sequence of a protein homologous to a TFSRP, which include fewer amino acids than a full length TFSRP or the full length protein which is homologous to a TFSRP, and exhibit at least one activity of a TFSRP. Typically, biologically active portions (e.g., peptides which are, for example, 5, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) comprise a domain or motif with at least one activity of a TFSRP. Moreover, other biologically active portions in which other regions of the protein

are deleted, can be prepared by recombinant techniques and evaluated for one or more of the activities described herein. Preferably, the biologically active portions of a TFSRP include one or more selected domains/motifs or portions thereof having biological activity.

[0052] The invention also provides TFSRP chimeric or fusion proteins. As used herein, a TFSRP "chimeric protein" or "fusion protein" comprises a TFSRP polypeptide operatively linked to a non-TFSRP polypeptide. A TFSRP polypeptide refers to a polypeptide having an amino acid sequence corresponding to a TFSRP, whereas a non-TFSRP polypeptide refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the TFSRP, e.g., a protein that is different from the TFSRP and is derived from the same or a different organism. Within the fusion protein, the term "operatively linked" is intended to indicate that the TFSRP polypeptide and the non-TFSRP polypeptide are fused to each other so that both sequences fulfill the proposed function attributed to the sequence used. The non-TFSRP polypeptide can be fused to the N-terminus or C-terminus of the TFSRP polypeptide. For example, in one embodiment, the fusion protein is a GST-TFSRP fusion protein in which the TFSRP sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant TFSRPs. In another embodiment, the fusion protein is a TFSRP containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a TFSRP can be increased through use of a heterologous signal sequence.

[0053] Preferably, a TFSRP chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see, for example, Current Protocols in Molecular Biology, Eds. Ausubel et al. John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a

GST polypeptide). A TFSRP encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the TFSRP .

[0054] In addition to fragments and fusion proteins of the TFSRPs described herein, the present invention includes homologs and analogs of naturally occurring TFSRPs and TFSRP encoding nucleic acids in a plant. "Homologs" are defined herein as two nucleic acids or proteins that have similar, or "homologous", nucleotide or amino acid sequences, respectively. Homologs include allelic variants, orthologs, paralogs, agonists and antagonists of TFSRPs as defined hereafter. The term "homolog" further encompasses nucleic acid molecules that differ from one of the nucleotide sequences shown in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16 (and portions thereof) due to degeneracy of the genetic code and thus encode the same TFSRP as that encoded by the nucleotide sequences shown in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16. As used herein a "naturally occurring" TFSRP refers to a TFSRP amino acid sequence that occurs in nature. Preferably, a naturally occurring TFSRP comprises an amino acid sequence selected from the group consisting of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 and SEQ ID NO:24.

[0055] An agonist of the TFSRP can retain substantially the same, or a subset, of the biological activities of the TFSRP. An antagonist of the TFSRP can inhibit one or more of the activities of the naturally occurring form of the TFSRP. For example, the TFSRP antagonist can competitively bind to a downstream or upstream member of the cell membrane component metabolic cascade that includes the TFSRP, or bind to a TFSRP that mediates transport of compounds across such membranes, thereby preventing translocation from taking place.

[0056] Nucleic acid molecules corresponding to natural allelic variants and analogs, orthologs and paralogs of a TFSRP cDNA can be isolated based on their identity to the *Physcomitrella patens* TFSRP nucleic acids described herein using TFSRP cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. In an alternative embodiment, homologs of the TFSRP can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the TFSRP for TFSRP agonist or antagonist activity. In one embodiment, a variegated library of TFSRP variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of TFSRP variants can be

produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential TFSRP sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of TFSRP sequences therein. There are a variety of methods that can be used to produce libraries of potential TFSRP homologs from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene is then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential TFSRP sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang, S.A., 1983 Tetrahedron 39:3; Itakura et al., 1984 Annu. Rev. Biochem. 53:323; Itakura et al., 1984 Science 198:1056; Ike et al., 1983 Nucleic Acid Res. 11:477).

[0057] In addition, libraries of fragments of the TFSRP coding regions can be used to generate a variegated population of TFSRP fragments for screening and subsequent selection of homologs of a TFSRP. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a TFSRP coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA, which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the TFSRP.

[0058] Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of TFSRP homologs. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the

screening assays to identify TFSRP homologs (Arkin and Yourvan, 1992 PNAS 89:7811-7815; Delgrave et al., 1993 Protein Engineering 6(3):327-331). In another embodiment, cell based assays can be exploited to analyze a variegated TFSRP library, using methods well known in the art. The present invention further provides a method of identifying a novel TFSRP, comprising (a) raising a specific antibody response to a TFSRP, or a fragment thereof, as described above; (b) screening putative TFSRP material with the antibody, wherein specific binding of the antibody to the material indicates the presence of a potentially novel TFSRP; and (c) analyzing the bound material in comparison to known TFSRP, to determine its novelty.

[0059] To determine the percent homology of two amino acid sequences (e.g., one of the sequences of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 and SEQ ID NO:24 and a mutant form thereof), the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues at corresponding amino acid positions are then compared. When a position in one sequence (e.g., one of the sequences of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 and SEQ ID NO:24) is occupied by the same amino acid residue as the corresponding position in the other sequence (e.g., a mutant form of the sequence selected from the polypeptide of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 and SEQ ID NO:24), then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The same type of comparison can be made between two nucleic acid sequences.

[0060] The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % homology = numbers of identical positions/total numbers of positions x 100). Preferably, the amino acid sequences included in the present invention are at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence shown in SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 or SEQ ID NO:24. In yet another embodiment, at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an

entire amino acid sequence encoded by a nucleic acid sequence shown in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16. In other embodiments, the preferable length of sequence comparison for proteins is at least 15 amino acid residues, more preferably at least 25 amino acid residues, and most preferably at least 35 amino acid residues.

[0061] In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which is at least about 50-60%, preferably at least about 60-70%, more preferably at least about 70-80%, 80-90%, or 90-95%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence shown in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16, or a portion thereof. The preferable length of sequence comparison for nucleic acids is at least 75 nucleotides, more preferably at least 100 nucleotides and most preferably the entire coding region.

[0062] It is also preferable that the homologous nucleic acid molecule of the invention encodes a protein or portion thereof which includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 or SEQ ID NO:24 such that the protein or portion thereof maintains the same or a similar function as the amino acid sequence to which it is compared. Functions of the TFSRP amino acid sequences of the present invention include the ability to participate in a stress tolerance response in a plant, or more particularly, to participate in the transcription of a protein involved in a stress tolerance response in a *Physcomitrella patens* plant. Examples of such activities are described in Table 1.

[0063] In addition to the above described methods, a determination of the percent homology between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990 Proc. Natl. Acad. Sci. USA 90:5873-5877). Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990 J. Mol. Biol. 215:403-410).

[0064] BLAST nucleic acid searches can be performed with the NBLAST program, score=100, wordlength=12 to obtain nucleic acid sequences homologous to the TFSRP nucleic acid molecules of the invention. Additionally, BLAST protein searches can be performed with the XBLAST program, score=50, wordlength=3 to obtain amino acid sequences homologous to TFSRPs of the present invention. To obtain gapped alignments for

comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997 Nucleic Acids Res. 25:3389-3402). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. Another preferred non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller (CABIOS 1989). Such an algorithm is incorporated into the ALIGN program (version 2.0) that is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4 can be used to obtain amino acid sequences homologous to the TFSRPs of the present invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997 Nucleic Acids Res. 25:3389-3402). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. Another preferred non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller (CABIOS 1989). Such an algorithm is incorporated into the ALIGN program (version 2.0) that is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4 can be used.

[0065] Finally, homology between nucleic acid sequences can also be determined using hybridization techniques known to those of skill in the art. Accordingly, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to one of the nucleotide sequences shown in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16, or a portion thereof. More particularly, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising a nucleotide sequence of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16. In other embodiments, the nucleic acid is at least 30, 50, 100, 250 or more nucleotides in length.

[0066] As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even more preferably at least about 75% or more homologous to each other typically

remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, 6.3.1-6.3.6, John Wiley & Sons, N.Y. (1989). A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16 corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein). In one embodiment, the nucleic acid encodes a naturally occurring *Physcomitrella patens* TFSRP.

[0067] Using the above-described methods, and others known to those of skill in the art, one of ordinary skill in the art can isolate homologs of the TFSRPs comprising amino acid sequences shown in SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 or SEQ ID NO:24 and the TFSRP nucleic acids comprising the nucleotide sequences shown in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16. One subset of these homologs are allelic variants. As used herein, the term "allelic variant" refers to a nucleotide sequence containing polymorphisms that lead to changes in the amino acid sequences of a TFSRP and that exist within a natural population (e.g., a plant species or variety). Such natural allelic variations can typically result in 1-5% variance in a TFSRP nucleic acid. Allelic variants can be identified by sequencing the nucleic acid sequence of interest in a number of different plants, which can be readily carried out by using hybridization probes to identify the same TFSRP genetic locus in those plants. Any and all such nucleic acid variations and resulting amino acid polymorphisms or variations in a TFSRP that are the result of natural allelic variation and that do not alter the functional activity of a TFSRP, are intended to be within the scope of the invention.

[0068] Moreover, nucleic acid molecules encoding TFSRPs from the same or other species such as TFSRP analogs, orthologs and paralogs, are intended to be within the scope of the present invention. As used herein, the term "analogs" refers to two nucleic acids that have the same or similar function, but that have evolved separately in unrelated organisms. As used herein, the term "orthologs" refers to two nucleic acids from different species, but that have evolved from a common ancestral gene by speciation. Normally, orthologs encode

proteins having the same or similar functions. As also used herein, the term "paralogs" refers to two nucleic acids that are related by duplication within a genome. Paralogs usually have different functions, but these functions may be related (Tatusov, R.L. et al. 1997 Science 278(5338):631-637). Analogs, orthologs and paralogs of a naturally occurring TFSRP can differ from the naturally occurring TFSRP by post-translational modifications, by amino acid sequence differences, or by both. Post-translational modifications include *in vivo* and *in vitro* chemical derivatization of polypeptides, e.g., acetylation, carboxylation, phosphorylation, or glycosylation, and such modifications may occur during polypeptide synthesis or processing or following treatment with isolated modifying enzymes. In particular, orthologs of the invention will generally exhibit at least 80-85%, more preferably 90%, and most preferably 95%, 96%, 97%, 98% or even 99% identity or homology with all or part of a naturally occurring TFSRP amino acid sequence and will exhibit a function similar to a TFSRP. Orthologs of the present invention are also preferably capable of participating in the stress response in plants. In one embodiment, the TFSRP orthologs maintain the ability to participate in the metabolism of compounds necessary for the construction of cellular membranes in *Physcomitrella patens*, or in the transport of molecules across these membranes.

[0069] In addition to naturally-occurring variants of a TFSRP sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into a nucleotide sequence, such as the sequences of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16, thereby leading to changes in the amino acid sequence of the encoded TFSRP, without altering the functional ability of the TFSRP. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the proteins including a sequence of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the TFSRPs without altering the activity of said TFSRP, whereas an "essential" amino acid residue is required for TFSRP activity. Other amino acid residues, however, (e.g., those that are not conserved or only semi-conserved in the domain having TFSRP activity) may not be essential for activity and thus are likely to be amenable to alteration without altering TFSRP activity.

[0070] Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding TFSRPs that contain changes in amino acid residues that are not

essential for TFSRP activity. Such TFSRPs differ in amino acid sequence from a sequence contained in SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 or SEQ ID NO:24, yet retain at least one of the TFSRP activities described herein. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50% homologous to an amino acid sequence of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 or SEQ ID NO:24. Preferably, the protein encoded by the nucleic acid molecule is at least about 50-60% homologous to one of the sequences of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 or SEQ ID NO:24, more preferably at least about 60-70% homologous to one of the sequences of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 or SEQ ID NO:24, even more preferably at least about 70-80%, 80-90%, 90-95% homologous to one of the sequences of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 or SEQ ID NO:24, and most preferably at least about 96%, 97%, 98%, or 99% homologous to one of the sequences of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 or SEQ ID NO:24. The preferred TFSRP homologs of the present invention are preferably capable of participating in the stress tolerance response in a plant, or more particularly, participating in the transcription of a protein involved in a stress tolerance response in a *Physcomitrella patens* plant, or have one or more activities set forth in Table 1.

[0071] An isolated nucleic acid molecule encoding a TFSRP homologous to a protein sequence of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23 or SEQ ID NO:24 can be created by introducing one or more nucleotide substitutions, additions or deletions into a nucleotide sequence of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into one of the sequences of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A

"conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain.

[0072] Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a TFSRP is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a TFSRP coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for a TFSRP activity described herein to identify mutants that retain TFSRP activity. Following mutagenesis of one of the sequences of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16, the encoded protein can be expressed recombinantly and the activity of the protein can be determined by analyzing the stress tolerance of a plant expressing the protein as described in Example 7.

[0073] In addition to the nucleic acid molecules encoding the TFSRPs described above, another aspect of the invention pertains to isolated nucleic acid molecules that are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire TFSRP coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a TFSRP. The term "coding region" refers to the region of the nucleotide sequence comprising codons that are translated into amino acid residues (e.g., the entire coding region of ,,,,, comprises nucleotides 1 to ....). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding a TFSRP. The term "noncoding region" refers to 5' and 3' sequences that flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

[0074] In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of one of the nucleotide sequences shown in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16, or a portion thereof. A nucleic acid molecule that is complementary to one of the nucleotide sequences shown in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16 is one which is sufficiently complementary to one of the nucleotide sequences shown in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16 such that it can hybridize to one of the nucleotide sequences shown in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16, thereby forming a stable duplex.

[0075] Given the coding strand sequences encoding the TFSRPs disclosed herein (e.g., the sequences set forth in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of TFSRP mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of TFSRP mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of TFSRP mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length.

[0076] An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-

methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[0077] The antisense nucleic acid molecules of the invention are typically administered to a cell or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a TFSRP to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. The antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecule to a peptide or an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong prokaryotic, viral, or eukaryotic (including plant) promoter are preferred.

[0078] In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier et al., 1987 Nucleic Acids. Res. 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al., 1987 Nucleic Acids Res. 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al., 1987 FEBS Lett. 215:327-330).

[0079] In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are

capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes described in Haselhoff and Gerlach, 1988 Nature 334:585-591) can be used to catalytically cleave TFSRP mRNA transcripts to thereby inhibit translation of TFSRP mRNA. A ribozyme having specificity for a TFSRP-encoding nucleic acid can be designed based upon the nucleotide sequence of a TFSRP cDNA, as disclosed herein (i.e., SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16) or on the basis of a heterologous sequence to be isolated according to methods taught in this invention. For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a TFSRP-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071 and Cech et al. U.S. Patent No. 5,116,742. Alternatively, TFSRP mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W., 1993 Science 261:1411-1418.

[0080] Alternatively, TFSRP gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of a TFSRP nucleotide sequence (e.g., a TFSRP promoter and/or enhancer) to form triple helical structures that prevent transcription of a TFSRP gene in target cells. See generally, Helene, C., 1991 Anticancer Drug Des. 6(6):569-84; Helene, C. et al., 1992 Ann. N.Y. Acad. Sci. 660:27-36; and Maher, L.J., 1992 Bioassays 14(12):807-15.

[0081] In addition to the TFSRP nucleic acids and proteins described above, the present invention encompasses these nucleic acids and proteins attached to a moiety. These moieties include, but are not limited to, detection moieties, hybridization moieties, purification moieties, delivery moieties, reaction moieties, binding moieties, and the like. One typical group of nucleic acids attached to a moiety are probes and primers. The probe/primer typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50 or 75 consecutive nucleotides of a sense strand of one of the sequences set forth in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16, an anti-sense sequence of one of the sequences set forth in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 and SEQ ID NO:16, or naturally occurring mutants thereof. Primers based on a nucleotide sequence of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID

NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16 can be used in PCR reactions to clone TFSRP homologs. Probes based on the TFSRP nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g. the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a genomic marker test kit for identifying cells which express a TFSRP , such as by measuring a level of a TFSRP-encoding nucleic acid, in a sample of cells, e.g., detecting TFSRP mRNA levels or determining whether a genomic TFSRP gene has been mutated or deleted.

[0082] In particular, a useful method to ascertain the level of transcription of the gene (an indicator of the amount of mRNA available for translation to the gene product) is to perform a Northern blot (for reference see, for example, Ausubel et al., 1988 Current Protocols in Molecular Biology, Wiley: New York). This information at least partially demonstrates the degree of transcription of the transformed gene. Total cellular RNA can be prepared from cells, tissues or organs by several methods, all well-known in the art, such as that described in Bormann, E.R. et al., 1992 Mol. Microbiol. 6:317-326. To assess the presence or relative quantity of protein translated from this mRNA, standard techniques, such as a Western blot, may be employed. These techniques are well known to one of ordinary skill in the art. (See, for example, Ausubel et al., 1988 Current Protocols in Molecular Biology, Wiley: New York).

[0083] The invention further provides an isolated recombinant expression vector comprising a TFSRP nucleic acid as described above, wherein expression of the vector in a host cell results in increased tolerance to environmental stress as compared to a wild type variety of the host cell. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general,

expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

[0084] The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/ translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990) or see: Gruber and Crosby, in: Methods in Plant Molecular Biology and Biotechnology, eds. Glick and Thompson, Chapter 7, 89-108, CRC Press: Boca Raton, Florida, including the references therein. Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cells and those that direct expression of the nucleotide sequence only in certain host cells or under certain conditions. It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., TFSRPs, mutant forms of TFSRPs, fusion proteins, etc.).

[0085] The recombinant expression vectors of the invention can be designed for expression of TFSRPs in prokaryotic or eukaryotic cells. For example, TFSRP genes can be expressed in bacterial cells such as *C. glutamicum*, insect cells (using baculovirus expression vectors), yeast and other fungal cells (see Romanos, M.A. et al., 1992 Foreign gene expression in yeast: a review, Yeast 8:423-488; van den Hondel, C.A.M.J.J. et al., 1991

Heterologous gene expression in filamentous fungi, in: More Gene Manipulations in Fungi, J.W. Bennet & L.L. Lasure, eds., p. 396-428: Academic Press: San Diego; and van den Hondel, C.A.M.J.J. & Punt, P.J., 1991 Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, Peberdy, J.F. et al., eds., p. 1-28, Cambridge University Press: Cambridge), algae (Falciatore et al., 1999 Marine Biotechnology 1(3):239-251), ciliates of the types: Holotrichia, Peritrichia, Spirotrichia, Suctoria, Tetrahymena, Paramecium, Colpidium, Glaucoma, Platyophrya, Potomacus, Pseudocohnilembus, Euplotes, Engelmaniella, and Stylonychia, especially of the genus Stylonychia lemnae with vectors following a transformation method as described in WO 98/01572 and multicellular plant cells (see Schmidt, R. and Willmitzer, L., 1988 High efficiency *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis thaliana* leaf and cotyledon explants, Plant Cell Rep. 583-586); Plant Molecular Biology and Biotechnology, C Press, Boca Raton, Florida, chapter 6/7, S.71-119 (1993); F.F. White, B. Jenes et al., Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, eds. Kung und R. Wu, 128-43, Academic Press: 1993; Potrykus, 1991 Annu. Rev. Plant Physiol. Plant Molec. Biol. 42:205-225 and references cited therein) or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press: San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[0086] Expression of proteins in prokaryotes is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein but also to the C-terminus or fused within suitable regions in the proteins. Such fusion vectors typically serve three purposes: 1) to increase expression of a recombinant protein; 2) to increase the solubility of a recombinant protein; and 3) to aid in the purification of a recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

[0087] Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S., 1988 Gene 67:31-40), pMAL (New England Biolabs,

Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In one embodiment, the coding sequence of the TFSRP is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from the N-terminus to the C-terminus, GST-thrombin cleavage site-X protein. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin. Recombinant TFSRP unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

[0088] Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., 1988 Gene 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a co-expressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

[0089] One strategy to maximize recombinant protein expression is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the bacterium chosen for expression, such as *C. glutamicum* (Wada et al., 1992 Nucleic Acids Res. 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

[0090] In another embodiment, the TFSRP expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, et al., 1987 Embo J. 6:229-234), pMFa (Kurjan and Herskowitz, 1982 Cell 30:933-943), pJRY88 (Schultz et al., 1987 Gene 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Vectors and methods for the construction of vectors appropriate for use in other fungi, such as the filamentous fungi, include those detailed in: van den Hondel, C.A.M.J.J. & Punt, P.J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: Applied Molecular Genetics of Fungi, J.F. Peberdy, et al., eds., p. 1-28, Cambridge University Press: Cambridge.

[0091] Alternatively, the TFSRPs of the invention can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al., 1983 Mol. Cell Biol. 3:2156-2165) and the pVL series (Lucklow and Summers, 1989 Virology 170:31-39).

[0092] In yet another embodiment, a TFSRP nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B., 1987 Nature 329:840) and pMT2PC (Kaufman et al., 1987 EMBO J. 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

[0093] In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al., 1987 Genes Dev. 1:268-277), lymphoid-specific promoters (Calame and Eaton, 1988 Adv. Immunol. 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989 EMBO J. 8:729-733) and immunoglobulins (Banerji et al., 1983 Cell 33:729-740; Queen and Baltimore, 1983 Cell 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989 *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund et al., 1985 Science 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example, the murine hox promoters (Kessel and Gruss, 1990 Science 249:374-379) and the fetoprotein promoter (Campes and Tilghman, 1989 Genes Dev. 3:537-546).

[0094] In another embodiment, the TFSRPs of the invention may be expressed in unicellular plant cells (such as algae) (see Falciatore et al., 1999 Marine Biotechnology 1(3):239-251 and references therein) and plant cells from higher plants (e.g., the spermatophytes, such as crop plants). Examples of plant expression vectors include those

detailed in: Becker, D., Kemper, E., Schell, J. and Masterson, R., 1992 New plant binary vectors with selectable markers located proximal to the left border, Plant Mol. Biol. 20: 1195-1197; and Bevan, M.W., 1984 Binary *Agrobacterium* vectors for plant transformation, Nucl. Acid. Res. 12:8711-8721; Vectors for Gene Transfer in Higher Plants; in: Transgenic Plants, Vol. 1, Engineering and Utilization, eds.: Kung and R. Wu, Academic Press, 1993, S. 15-38.

[0095] A plant expression cassette preferably contains regulatory sequences capable of driving gene expression in plant cells and operably linked so that each sequence can fulfill its function, for example, termination of transcription by polyadenylation signals. Preferred polyadenylation signals are those originating from *Agrobacterium tumefaciens* t-DNA such as the gene 3 known as octopine synthase of the Ti-plasmid pTiACH5 (Gielen et al., 1984 EMBO J. 3:835) or functional equivalents thereof but also all other terminators functionally active in plants are suitable.

[0096] As plant gene expression is very often not limited on transcriptional levels, a plant expression cassette preferably contains other operably linked sequences like translational enhancers such as the overdrive-sequence containing the 5′-untranslated leader sequence from tobacco mosaic virus enhancing the protein per RNA ratio (Gallie et al., 1987 Nucl. Acids Research 15:8693-8711).

[0097] Plant gene expression has to be operably linked to an appropriate promoter conferring gene expression in a timely, cell or tissue specific manner. Preferred are promoters driving constitutive expression (Benfey et al., 1989 EMBO J. 8:2195-2202) like those derived from plant viruses like the 35S CAMV (Franck et al., 1980 Cell 21:285-294), the 19S CaMV (see also U.S. Patent No. 5352605 and PCT Application No. WO 8402913) or plant promoters like those from Rubisco small subunit described in U.S. Patent No. 4,962,028.

[0098] Other preferred sequences for use in plant gene expression cassettes are targeting-sequences necessary to direct the gene product in its appropriate cell compartment (for review see Kermode, 1996 Crit. Rev. Plant Sci. 15(4):285-423 and references cited therein) such as the vacuole, the nucleus, all types of plastids like amyloplasts, chloroplasts, chromoplasts, the extracellular space, mitochondria, the endoplasmic reticulum, oil bodies, peroxisomes and other compartments of plant cells.

[0099] Plant gene expression can also be facilitated via an inducible promoter (for review see Gatz, 1997 Annu. Rev. Plant Physiol. Plant Mol. Biol. 48:89-108). Chemically inducible promoters are especially suitable if gene expression is wanted to occur in a time

specific manner. Examples of such promoters are a salicylic acid inducible promoter (PCT Application No. WO 95/19443), a tetracycline inducible promoter (Gatz et al., 1992 Plant J. 2:397-404) and an ethanol inducible promoter (PCT Application No. WO 93/21334).

[0100] Also, suitable promoters responding to biotic or abiotic stress conditions are those such as the pathogen inducible PRP1-gene promoter (Ward et al., 1993 Plant. Mol. Biol. 22:361-366), the heat inducible hsp80-promoter from tomato (U.S. Patent No. 5187267), cold inducible alpha-amylase promoter from potato (PCT Application No. WO 96/12814) or the wound-inducible pinII-promoter (European Patent No. 375091). For other examples of drought, cold, and salt-inducible promoters, such as the RD29A promoter, see Yamaguchi-Shinozalei et al. (1993 Mol. Gen. Genet. 236:331-340).

[0101] Especially preferred are those promoters that confer gene expression in specific tissues and organs, such as guard cells and the root hair cells. Suitable promoters include the napin-gene promoter from rapeseed (U.S. Patent No. 5,608,152), the USP-promoter from Vicia faba (Baeumlein et al., 1991 Mol Gen Genet. 225(3):459-67), the oleosin-promoter from *Arabidopsis* (PCT Application No. WO 98/45461), the phaseolin-promoter from Phaseolus vulgaris (U.S. Patent No. 5,504,200), the Bce4-promoter from Brassica (PCT Application No. WO 91/13980) or the legumin B4 promoter (LeB4; Baeumlein et al., 1992 Plant Journal, 2(2):233-9) as well as promoters conferring seed specific expression in monocot plants like maize, barley, wheat, rye, rice, etc. Suitable promoters to note are the lpt2 or lpt1-gene promoter from barley (PCT Application No. WO 95/15389 and PCT Application No. WO 95/23230) or those described in PCT Application No. WO 99/16890 (promoters from the barley hordein-gene, rice glutelin gene, rice oryzin gene, rice prolamin gene, wheat gliadin gene, wheat glutelin gene, maize zein gene, oat glutelin gene, Sorghum kasirin-gene and rye secalin gene).

[0102] Also especially suited are promoters that confer plastid-specific gene expression since plastids are the compartment where lipid biosynthesis occurs. Suitable promoters are the viral RNA-polymerase promoter described in PCT Application No. WO 95/16783 and PCT Application No. WO 97/06250 and the clpP-promoter from *Arabidopsis* described in PCT Application No. WO 99/46394.

[0103] The invention further provides a recombinant expression vector comprising a TFSRP DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to a TFSRP mRNA. Regulatory sequences operatively linked to a

nucleic acid molecule cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types. For instance, viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus wherein antisense nucleic acids are produced under the control of a high efficiency regulatory region. The activity of the regulatory region can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., Antisense RNA as a molecular tool for genetic analysis, Reviews - Trends in Genetics, Vol. 1(1) 1986 and Mol et al., 1990 FEBS Letters 268:427-430.

[0104] Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but they also apply to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[0105] A host cell can be any prokaryotic or eukaryotic cell. For example, a TFSRP can be expressed in bacterial cells such as *C. glutamicum*, insect cells, fungal cells or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells), algae, ciliates, plant cells, fungi or other microorganisms like *C. glutamicum*. Other suitable host cells are known to those skilled in the art.

[0106] Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation", "transfection", "conjugation" and "transduction" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, natural competence, chemical-mediated transfer and electroporation. Suitable methods for transforming or transfecting host cells including plant cells can be found in Sambrook, et al. (Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989) and other laboratory manuals such as Methods in Molecular Biology, 1995, Vol. 44, *Agrobacterium* protocols, ed: Gartland and Davey, Humana Press, Totowa, New Jersey.

As biotic and abiotic stress tolerance is a general trait wished to be inherited into a wide variety of plants like maize, wheat, rye, oat, triticale, rice, barley, soybean, peanut, cotton, rapeseed and canola, manihot, pepper, sunflower and tagetes, solanaceous plants like potato, tobacco, eggplant, and tomato, Vicia species, pea, alfalfa, bushy plants (coffee, cacao, tea), Salix species, trees (oil palm, coconut), perennial grasses and forage crops, these crop plants are also preferred target plants for a genetic engineering as one further embodiment of the present invention.

[0107] In particular, the invention provides a method of producing a transgenic plant with a TFSRP coding nucleic acid, wherein expression of the nucleic acid(s) in the plant results in increased tolerance to environmental stress as compared to a wild type variety of the plant comprising: (a) transforming a plant cell with an expression vector comprising a TFSRP nucleic acid, and (b) generating from the plant cell a transgenic plant with a increased tolerance to environmental stress as compared to a wild type variety of the plant. The invention also provides a method of increasing expression of a gene of interest within a host cell as compared to a wild type variety of the host cell, wherein the gene of interest is transcribed in response to a TFSRP, comprising: (a) transforming the host cell with an expression vector comprising a TFSRP coding nucleic acid, and (b) expressing the TFSRP within the host cell, thereby increasing the expression of the gene transcribed in response to the TFSRP, as compared to a wild type variety of the host cell.

[0108] For such plant transformation, binary vectors such as pBinAR can be used (Höfgen and Willmitzer, 1990 Plant Science 66:221-230). Construction of the binary vectors can be performed by ligation of the cDNA in sense or antisense orientation into the T-DNA. 5-prime to the cDNA a plant promoter activates transcription of the cDNA. A polyadenylation sequence is located 3-prime to the cDNA. Tissue-specific expression can be achieved by using a tissue specific promoter. For example, seed-specific expression can be achieved by cloning the napin or LeB4 or USP promoter 5-prime to the cDNA. Also, any other seed specific promoter element can be used. For constitutive expression within the whole plant, the CaMV 35S promoter can be used. The expressed protein can be targeted to a cellular compartment using a signal peptide, for example for plastids, mitochondria or endoplasmic reticulum (Kermode, 1996 Crit. Rev. Plant Sci. 4 (15):285-423). The signal peptide is cloned 5-prime in frame to the cDNA to archive subcellular localization of the fusion protein. Additionally, promoters that are responsive to abiotic stresses can be used with, such as the *Arabidopsis* promoter RD29A, the nucleic acid sequences disclosed herein. One skilled in the art will recognize that the promoter used should be operatively linked to

the nucleic acid such that the promoter causes transcription of the nucleic acid which results in the synthesis of a mRNA which encodes a polypeptide. Alternatively, the RNA can be an antisense RNA for use in affecting subsequent expression of the same or another gene or genes.

[0109] Alternate methods of transfection include the direct transfer of DNA into developing flowers via electroporation or *Agrobacterium* mediated gene transfer. *Agrobacterium* mediated plant transformation can be performed using for example the GV3101(pMP90) (Koncz and Schell, 1986 Mol. Gen. Genet. 204:383-396) or LBA4404 (Clontech) *Agrobacterium tumefaciens* strain. Transformation can be performed by standard transformation and regeneration techniques (Deblaere et al., 1994 Nucl. Acids. Res. 13:4777-4788; Gelvin, Stanton B. and Schilperoort, Robert A, Plant Molecular Biology Manual, 2nd Ed. - Dordrecht : Kluwer Academic Publ., 1995. - in Sect., Ringbuc Zentrale Signatur: BT11-P ISBN 0-7923-2731-4; Glick, Bernard R.; Thompson, John E., Methods in Plant Molecular Biology and Biotechnology, Boca Raton : CRC Press, 1993. - 360 S.,ISBN 0-8493-5164-2). For example, rapeseed can be transformed via cotyledon or hypocotyl transformation (Moloney et al., 1989 Plant cell Report 8:238-242; De Block et al., 1989 Plant Physiol. 91:694-701). Use of antibiotica for *Agrobacterium* and plant selection depends on the binary vector and the *Agrobacterium* strain used for transformation. Rapeseed selection is normally performed using kanamycin as selectable plant marker. *Agrobacterium* mediated gene transfer to flax can be performed using, for example, a technique described by Mlynarova et al., 1994 Plant Cell Report 13:282-285. Additionally, transformation of soybean can be performed using for example a technique described in European Patent No. 0424 047, U.S. Patent No. 5,322,783, European Patent No. 0397 687, U.S. Patent No. 5,376,543 or U.S. Patent No. 5,169,770. Transformation of maize can be achieved by particle bombardment, polyethylene glycol mediated DNA uptake or via the silicon carbide fiber technique. (See, for example, Freeling and Walbot "The maize handbook" Springer Verlag: New York (1993) ISBN 3-540-97826-7). A specific example of maize transformation is found in U.S. Patent No. 5,990,387 and a specific example of wheat transformation can be found in PCT Application No. WO 93/07256.

[0110] For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those

which confer resistance to drugs, such as G418, hygromycin and methotrexate or in plants that confer resistance towards a herbicide such as glyphosate or glufosinate. Nucleic acid molecules encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a TFSRP or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid molecule can be identified by, for example, drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

[0111] To create a homologous recombinant microorganism, a vector is prepared which contains at least a portion of a TFSRP gene into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the TFSRP gene. Preferably, the TFSRP gene is a *Physcomitrella patens* TFSRP gene, but it can be a homolog from a related plant or even from a mammalian, yeast, or insect source. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous TFSRP gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a knock-out vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous TFSRP gene is mutated or otherwise altered but still encodes a functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous TFSRP). To create a point mutation via homologous recombination, DNA-RNA hybrids can be used in a technique known as chimeraplasty (Cole-Strauss et al., 1999 Nucleic Acids Research 27(5):1323-1330 and Kmiec, 1999 Gene therapy American Scientist. 87(3):240-247). Homologous recombination procedures in *Physcomitrella patens* are also well known in the art and are contemplated for use herein.

[0112] Whereas in the homologous recombination vector, the altered portion of the TFSRP gene is flanked at its 5' and 3' ends by an additional nucleic acid molecule of the TFSRP gene to allow for homologous recombination to occur between the exogenous TFSRP gene carried by the vector and an endogenous TFSRP gene, in a microorganism or plant. The additional flanking TFSRP nucleic acid molecule is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several hundreds of base pairs up to kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas, K.R., and Capecchi, M.R., 1987 Cell 51:503 for a description of homologous recombination vectors or Strepp et al., 1998 PNAS, 95 (8):4368-4373 for cDNA based recombination in *Physcomitrella patens*). The vector is introduced into a microorganism or plant cell (e.g., via polyethylene glycol mediated DNA), and cells in which

the introduced TFSRP gene has homologously recombined with the endogenous TFSRP gene are selected using art-known techniques.

[0113] In another embodiment, recombinant microorganisms can be produced that contain selected systems which allow for regulated expression of the introduced gene. For example, inclusion of a TFSRP gene on a vector placing it under control of the lac operon permits expression of the TFSRP gene only in the presence of IPTG. Such regulatory systems are well known in the art.

[0114] A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) a TFSRP. Accordingly, the invention further provides methods for producing TFSRPs using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a TFSRP has been introduced, or into which genome has been introduced a gene encoding a wild-type or altered TFSRP) in a suitable medium until TFSRP is produced. In another embodiment, the method further comprises isolating TFSRPs from the medium or the host cell.

[0115] Another aspect of the invention pertains to isolated TFSRPs, and biologically active portions thereof. An "isolated" or "purified" protein or biologically active portion thereof is free of some of the cellular material when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of TFSRP in which the protein is separated from some of the cellular components of the cells in which it is naturally or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of a TFSRP having less than about 30% (by dry weight) of non-TFSRP material (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-TFSRP material, still more preferably less than about 10% of non-TFSRP material, and most preferably less than about 5% non-TFSRP material.

[0116] When the TFSRP or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The language "substantially free of chemical precursors or other chemicals" includes preparations of TFSRP in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of a TFSRP having less than about 30%

(by dry weight) of chemical precursors or non-TFSRP chemicals, more preferably less than about 20% chemical precursors or non-TFSRP chemicals, still more preferably less than about 10% chemical precursors or non-TFSRP chemicals, and most preferably less than about 5% chemical precursors or non-TFSRP chemicals. In preferred embodiments, isolated proteins, or biologically active portions thereof, lack contaminating proteins from the same organism from which the TFSRP is derived. Typically, such proteins are produced by recombinant expression of, for example, a *Physcomitrella patens* TFSRP in plants other than *Physcomitrella patens* or microorganisms such as *C. glutamicum,* ciliates, algae or fungi.

[0117] The nucleic acid molecules, proteins, protein homologs, fusion proteins, primers, vectors, and host cells described herein can be used in one or more of the following methods: identification of *Physcomitrella patens* and related organisms; mapping of genomes of organisms related to *Physcomitrella patens*; identification and localization of *Physcomitrella patens* sequences of interest; evolutionary studies; determination of TFSRP regions required for function; modulation of a TFSRP activity; modulation of the metabolism of one or more cell functions; modulation of the transmembrane transport of one or more compounds; and modulation of stress resistance.

[0118] The moss *Physcomitrella patens* represents one member of the mosses. It is related to other mosses such as *Ceratodon purpureus* which is capable of growth in the absence of light. Mosses like *Ceratodon* and *Physcomitrella* share a high degree of homology on the DNA sequence and polypeptide level allowing the use of heterologous screening of DNA molecules with probes evolving from other mosses or organisms, thus enabling the derivation of a consensus sequence suitable for heterologous screening or functional annotation and prediction of gene functions in third species. The ability to identify such functions can therefore have significant relevance, e.g., prediction of substrate specificity of enzymes. Further, these nucleic acid molecules may serve as reference points for the mapping of moss genomes, or of genomes of related organisms.

[0119] The TFSRP nucleic acid molecules of the invention have a variety of uses. Most importantly, the nucleic acid and amino acid sequences of the present invention can be used to transform plants, thereby inducing tolerance to stresses such as drought, high salinity and cold. The present invention therefore provides a transgenic plant transformed by a TFSRP nucleic acid, wherein expression of the nucleic acid sequence in the plant results in increased tolerance to environmental stress as compared to a wild type variety of the plant. The transgenic plant can be a monocot or a dicot. The invention further provides that the transgenic plant can be selected from maize, wheat, rye, oat, triticale, rice, barley, soybean,

peanut, cotton, rapeseed, canola, manihot, pepper, sunflower, tagetes, solanaceous plants, potato, tobacco, eggplant, tomato, Vicia species, pea, alfalfa, coffee, cacao, tea, Salix species, oil palm, coconut, perennial grass and forage crops, for example.

[0120] In particular, the present invention describes using the expression of APS-2, ZF-2, ZF-3, ZF-4, ZF-5, MYB-1, CABF-3 and SFL-1 of *Physcomitrella patens* to engineer drought-tolerant, salt-tolerant and/or cold-tolerant plants. This strategy has herein been demonstrated for *Arabidopsis thaliana*, Rapeseed/Canola, soybeans, corn and wheat but its application is not restricted to these plants. Accordingly, the invention provides a transgenic plant containing a TFSRP selected from APS-2 (SEQ ID NO:17), ZF-2 (SEQ ID NO:18), ZF-3 (SEQ ID NO:19), ZF-4 (SEQ ID NO:20), ZF-5 (SEQ ID NO:21), MYB-1 (SEQ ID NO:22), CABF-3 (SEQ ID NO:23) and SFL-1 (SEQ ID NO:24), wherein the environmental stress is drought, increased salt or decreased or increased temperature. In preferred embodiments, the environmental stress is drought or decreased temperature

[0121] The present invention also provides methods of modifying stress tolerance of a plant comprising, modifying the expression of a TFSRP in the plant. The invention provides that this method can be performed such that the stress tolerance is either increased or decreased. In particular, the present invention provides methods of producing a transgenic plant having an increased tolerance to environmental stress as compared to a wild type variety of the plant comprising increasing expression of a TFSRP in a plant.

[0122] The methods of increasing expression of TFSRPs can be used wherein the plant is either transgenic or not transgenic. In cases when the plant is transgenic, the plant can be transformed with a vector containing any of the above described TFSRP coding nucleic acids, or the plant can be transformed with a promoter that directs expression of native TFSRP in the plant, for example. The invention provides that such a promoter can be tissue specific. Furthermore, such a promoter can be developmentally regulated. Alternatively, non-transgenic plants can have native TFSRP expression modified by inducing a native promoter.

[0123] The expression of APS-2 (SEQ ID NO:17), ZF-2 (SEQ ID NO:18), ZF-3 (SEQ ID NO:19), ZF-4 (SEQ ID NO:20), ZF-5 (SEQ ID NO:21), MYB-1 (SEQ ID NO:22), CABF-3 (SEQ ID NO:23) or SFL-1 (SEQ ID NO:24) in target plants can be accomplished by, but is not limited to, one of the following examples: (a) constitutive promoter, (b) stress-inducible promoter, (c) chemical-induced promoter, and (d) engineered promoter over-expression with for example zinc-finger derived transcription factors (Greisman and Pabo, 1997 Science 275:657). The later case involves identification of the APS-2 (SEQ ID NO:17),

ZF-2 (SEQ ID NO:18), ZF-3 (SEQ ID NO:19), ZF-4 (SEQ ID NO:20), ZF-5 (SEQ ID NO:21), MYB-1 (SEQ ID NO:22), CABF-3 (SEQ ID NO:23) or SFL-1 (SEQ ID NO:24) homologs in the target plant as well as from its promoter. Zinc-finger-containing recombinant transcription factors are engineered to specifically interact with the APS-2 (SEQ ID NO:17), ZF-2 (SEQ ID NO:18), ZF-3 (SEQ ID NO:19), ZF-4 (SEQ ID NO:20), ZF-5 (SEQ ID NO:21), MYB-1 (SEQ ID NO:22), CABF-3 (SEQ ID NO:23) or SFL-1 (SEQ ID NO:24) homolog and transcription of the corresponding gene is activated.

[0124] In addition to introducing the TFSRP nucleic acid sequences into transgenic plants, these sequences can also be used to identify an organism as being *Physcomitrella patens* or a close relative thereof. Also, they may be used to identify the presence of *Physcomitrella patens* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *Physcomitrella patens* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *Physcomitrella patens* gene which is unique to this organism, one can ascertain whether this organism is present.

[0125] Further, the nucleic acid and protein molecules of the invention may serve as markers for specific regions of the genome. This has utility not only in the mapping of the genome, but also in functional studies of *Physcomitrella patens* proteins. For example, to identify the region of the genome to which a particular *Physcomitrella patens* DNA-binding protein binds, the *Physcomitrella patens* genome could be digested, and the fragments incubated with the DNA-binding protein. Those fragments that bind the protein may be additionally probed with the nucleic acid molecules of the invention, preferably with readily detectable labels. Binding of such a nucleic acid molecule to the genome fragment enables the localization of the fragment to the genome map of *Physcomitrella patens*, and, when performed multiple times with different enzymes, facilitates a rapid determination of the nucleic acid sequence to which the protein binds. Further, the nucleic acid molecules of the invention may be sufficiently homologous to the sequences of related species such that these nucleic acid molecules may serve as markers for the construction of a genomic map in related mosses.

[0126] The TFSRP nucleic acid molecules of the invention are also useful for evolutionary and protein structural studies. The metabolic and transport processes in which the molecules of the invention participate are utilized by a wide variety of prokaryotic and eukaryotic cells; by comparing the sequences of the nucleic acid molecules of the present

invention to those encoding similar enzymes from other organisms, the evolutionary relatedness of the organisms can be assessed. Similarly, such a comparison permits an assessment of which regions of the sequence are conserved and which are not, which may aid in determining those regions of the protein that are essential for the functioning of the enzyme. This type of determination is of value for protein engineering studies and may give an indication of what the protein can tolerate in terms of mutagenesis without losing function.

[0127] Manipulation of the TFSRP nucleic acid molecules of the invention may result in the production of TFSRPs having functional differences from the wild-type TFSRPs. These proteins may be improved in efficiency or activity, may be present in greater numbers in the cell than is usual, or may be decreased in efficiency or activity.

[0128] There are a number of mechanisms by which the alteration of a TFSRP of the invention may directly affect stress response and/or stress tolerance. In the case of plants expressing TFSRPs, increased transport can lead to improved salt and/or solute partitioning within the plant tissue and organs. By either increasing the number or the activity of transporter molecules which export ionic molecules from the cell, it may be possible to affect the salt tolerance of the cell.

[0129] The effect of the genetic modification in plants, *C. glutamicum*, fungi, algae, or ciliates on stress tolerance can be assessed by growing the modified microorganism or plant under less than suitable conditions and then analyzing the growth characteristics and/or metabolism of the plant. Such analysis techniques are well known to one skilled in the art, and include dry weight, wet weight, protein synthesis, carbohydrate synthesis, lipid synthesis, evapotranspiration rates, general plant and/or crop yield, flowering, reproduction, seed setting, root growth, respiration rates, photosynthesis rates, etc. (Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17; Rehm et al., 1993 Biotechnology, vol. 3, Chapter III: Product recovery and purification, page 469-714, VCH: Weinheim; Belter, P.A. et al., 1988 Bioseparations: downstream processing for biotechnology, John Wiley and Sons; Kennedy, J.F. and Cabral, J.M.S., 1992 Recovery processes for biological materials, John Wiley and Sons; Shaeiwitz, J.A. and Henry, J.D., 1988 Biochemical separations, in: Ulmann's Encyclopedia of Industrial Chemistry, vol. B3, Chapter 11, page 1-27, VCH: Weinheim; and Dechow, F.J. (1989) Separation and purification techniques in biotechnology, Noyes Publications).

[0130] For example, yeast expression vectors comprising the nucleic acids disclosed herein, or fragments thereof, can be constructed and transformed into *Saccharomyces cerevisiae* using standard protocols. The resulting transgenic cells can then be assayed for fail

or alteration of their tolerance to drought, salt, and temperature stress. Similarly, plant expression vectors comprising the nucleic acids disclosed herein, or fragments thereof, can be constructed and transformed into an appropriate plant cell such as *Arabidopsis*, soy, rape, maize, wheat, *Medicago truncatula*, etc., using standard protocols. The resulting transgenic cells and/or plants derived there from can then be assayed for fail or alteration of their tolerance to drought, salt, and temperature stress.

[0131] The engineering of one or more TFSRP genes of the invention may also result in TFSRPs having altered activities which indirectly impact the stress response and/or stress tolerance of algae, plants, ciliates or fungi or other microorganisms like *C. glutamicum*. For example, the normal biochemical processes of metabolism result in the production of a variety of products (e.g., hydrogen peroxide and other reactive oxygen species) which may actively interfere with these same metabolic processes (for example, peroxynitrite is known to nitrate tyrosine side chains, thereby inactivating some enzymes having tyrosine in the active site (Groves, J.T., 1999 Curr. Opin. Chem. Biol. 3(2):226-235). While these products are typically excreted, cells can be genetically altered to transport more products than is typical for a wild-type cell. By optimizing the activity of one or more TFSRPs of the invention which are involved in the export of specific molecules, such as salt molecules, it may be possible to improve the stress tolerance of the cell.

[0132] Additionally, the sequences disclosed herein, or fragments thereof, can be used to generate knockout mutations in the genomes of various organisms, such as bacteria, mammalian cells, yeast cells, and plant cells (Girke, T., 1998 The Plant Journal 15:39-48). The resultant knockout cells can then be evaluated for their ability or capacity to tolerate various stress conditions, their response to various stress conditions, and the effect on the phenotype and/or genotype of the mutation. For other methods of gene inactivation see U.S. Patent No. 6004804 "Non-Chimeric Mutational Vectors" and Puttaraju et al., 1999 Spliceosome-mediated RNA *trans*-splicing as a tool for gene therapy Nature Biotechnology 17:246-252.

[0133] The aforementioned mutagenesis strategies for TFSRPs resulting in increased stress resistance are not meant to be limiting; variations on these strategies will be readily apparent to one skilled in the art. Using such strategies, and incorporating the mechanisms disclosed herein, the nucleic acid and protein molecules of the invention may be utilized to generate algae, ciliates, plants, fungi or other microorganisms like *C. glutamicum* expressing mutated TFSRP nucleic acid and protein molecules such that the stress tolerance is improved.

[0134] The present invention also provides antibodies that specifically bind to a TFSRP, or a portion thereof, as encoded by a nucleic acid described herein. Antibodies can be made by many well-known methods (See, e.g. *Harlow and Lane*, "Antibodies; A Laboratory Manual" Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1988)). Briefly, purified antigen can be injected into an animal in an amount and in intervals sufficient to elicit an immune response. Antibodies can either be purified directly, or spleen cells can be obtained from the animal. The cells can then fused with an immortal cell line and screened for antibody secretion. The antibodies can be used to screen nucleic acid clone libraries for cells secreting the antigen. Those positive clones can then be sequenced. (See, for example, Kelly et al., 1992 Bio/Technology 10:163-167; Bebbington et al., 1992 Bio/Technology 10:169-175).

[0135] The phrases "selectively binds" and "specifically binds" with the polypeptide refer to a binding reaction that is determinative of the presence of the protein in a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bound to a particular protein do not bind in a significant amount to other proteins present in the sample. Selective binding of an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. A variety of immunoassay formats may be used to select antibodies that selectively bind with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select antibodies selectively immunoreactive with a protein. See *Harlow and Lane* "Antibodies, A Laboratory Manual" Cold Spring Harbor Publications, New York, (1988), for a description of immunoassay formats and conditions that could be used to determine selective binding.

[0136] In some instances, it is desirable to prepare monoclonal antibodies from various hosts. A description of techniques for preparing such monoclonal antibodies may be found in Stites et al., editors, "Basic and Clinical Immunology," (Lange Medical Publications, Los Altos, Calif., Fourth Edition) and references cited therein, and in Harlow and Lane ("Antibodies, A Laboratory Manual" Cold Spring Harbor Publications, New York, 1988).

[0137] Throughout this application, various publications are referenced. The disclosures of all of these publications and those references cited within those publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains.

[0138] It should also be understood that the foregoing relates to preferred embodiments of the present invention and that numerous changes may be made therein without departing from the scope of the invention. The invention is further illustrated by the following examples, which are not to be construed in any way as imposing limitations upon the scope thereof. On the contrary, it is to be clearly understood that resort may be had to various other embodiments, modifications, and equivalents thereof, which, after reading the description herein, may suggest themselves to those skilled in the art without departing from the spirit of the present invention and/or the scope of the appended claims.

EXAMPLES

Example 1

Growth of Physcomitrella patens cultures

[0139] For this study, plants of the species *Physcomitrella patens* (Hedw.) B.S.G. from the collection of the genetic studies section of the University of Hamburg were used. They originate from the strain 16/14 collected by H.L.K. Whitehouse in Gransden Wood, Huntingdonshire (England), which was subcultured from a spore by Engel (1968, Am. J. Bot. 55, 438-446). Proliferation of the plants was carried out by means of spores and by means of regeneration of the gametophytes. The protonema developed from the haploid spore as a chloroplast-rich chloronema and chloroplast-low caulonema, on which buds formed after approximately 12 days. These grew to give gametophores bearing antheridia and archegonia. After fertilization, the diploid sporophyte with a short seta and the spore capsule resulted, in which the meiospores matured.

[0140] Culturing was carried out in a climatic chamber at an air temperature of 25°C and light intensity of 55 micromols^{-1m2} (white light; Philips TL 65W/25 fluorescent tube) and a light/dark change of 16/8 hours. The moss was either modified in liquid culture using Knop medium according to Reski and Abel (1985, Planta 165:354-358) or cultured on Knop solid medium using 1% oxoid agar (Unipath, Basingstoke, England). The protonemas used for RNA and DNA isolation were cultured in aerated liquid cultures. The protonemas were comminuted every 9 days and transferred to fresh culture medium.

Example 2

Total DNA isolation from plants

[0141] The details for the isolation of total DNA relate to the working up of one gram fresh weight of plant material. The materials used include the following buffers: CTAB buffer: 2% (w/v) N-cethyl-N,N,N-trimethylammonium bromide (CTAB); 100 mM Tris HCl pH 8.0; 1.4 M NaCl; 20 mM EDTA; N-Laurylsarcosine buffer: 10% (w/v) N-laurylsarcosine; 100 mM Tris HCl pH 8.0; 20 mM EDTA.

[0142] The plant material was triturated under liquid nitrogen in a mortar to give a fine powder and transferred to 2 ml Eppendorf vessels. The frozen plant material was then covered with a layer of 1 ml of decomposition buffer (1 ml CTAB buffer, 100 µl of N-laurylsarcosine buffer, 20 µl of β-mercaptoethanol and 10 µl of proteinase K solution, 10 mg/ml) and incubated at 60°C for one hour with continuous shaking. The homogenate obtained was distributed into two Eppendorf vessels (2 ml) and extracted twice by shaking with the same volume of chloroform/isoamyl alcohol (24:1). For phase separation, centrifugation was carried out at 8000 x g and room temperature for 15 minutes in each case. The DNA was then precipitated at -70°C for 30 minutes using ice-cold isopropanol. The precipitated DNA was sedimented at 4°C and 10,000 g for 30 minutes and resuspended in 180 µl of TE buffer (Sambrook et al., 1989, Cold Spring Harbor Laboratory Press: ISBN 0-87969-309-6). For further purification, the DNA was treated with NaCl (1.2 M final concentration) and precipitated again at -70°C for 30 minutes using twice the volume of absolute ethanol. After a washing step with 70% ethanol, the DNA was dried and subsequently taken up in 50 µl of H_2O + RNAse (50 mg/ml final concentration). The DNA was dissolved overnight at 4°C and the RNAse digestion was subsequently carried out at 37°C for 1 hour. Storage of the DNA took place at 4°C.

Example 3

Isolation of total RNA and poly-(A)+ RNA and cDNA library construction from Physcomitrella patens

[0143] For the investigation of transcripts, both total RNA and poly-$(A)^+$ RNA were isolated. The total RNA was obtained from wild-type 9 day old protonemata following the GTC-method (Reski et al. 1994, Mol. Gen. Genet., 244:352-359). The Poly(A)+ RNA was isolated using Dyna Beads^R (Dynal, Oslo, Norway) following the instructions of the manufacturers protocol. After determination of the concentration of the RNA or of the

poly(A)+ RNA, the RNA was precipitated by addition of 1/10 volumes of 3 M sodium acetate pH 4.6 and 2 volumes of ethanol and stored at -70°C.

[0144] For cDNA library construction, first strand synthesis was achieved using Murine Leukemia Virus reverse transcriptase (Roche, Mannheim, Germany) and oligo-d(T)-primers, second strand synthesis by incubation with DNA polymerase I, Klenow enzyme and RNAseH digestion at 12°C (2 hours), 16°C (1 hour) and 22°C (1 hour). The reaction was stopped by incubation at 65°C (10 minutes) and subsequently transferred to ice. Double stranded DNA molecules were blunted by T4-DNA-polymerase (Roche, Mannheim) at 37°C (30 minutes). Nucleotides were removed by phenol/chloroform extraction and Sephadex G50 spin columns. EcoRI adapters (Pharmacia, Freiburg, Germany) were ligated to the cDNA ends by T4-DNA-ligase (Roche, 12°C, overnight) and phosphorylated by incubation with polynucleotide kinase (Roche, 37°C, 30 minutes). This mixture was subjected to separation on a low melting agarose gel. DNA molecules larger than 300 base pairs were eluted from the gel, phenol extracted, concentrated on Elutip-D-columns (Schleicher and Schuell, Dassel, Germany) and were ligated to vector arms and packed into lambda ZAPII phages or lambda ZAP-Express phages using the Gigapack Gold Kit (Stratagene, Amsterdam, Netherlands) using material and following the instructions of the manufacturer.

Example 4

Sequencing and function annotation of Physcomitrella patens ESTs

[0145] cDNA libraries as described in Example 3 were used for DNA sequencing according to standard methods, and in particular, by the chain termination method using the ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer, Weiterstadt, Germany). Random Sequencing was carried out subsequent to preparative plasmid recovery from cDNA libraries via *in vivo* mass excision, retransformation, and subsequent plating of DH10B on agar plates (material and protocol details from Stratagene, Amsterdam, Netherlands. Plasmid DNA was prepared from overnight grown E. coli cultures grown in Luria-Broth medium containing ampicillin (see Sambrook et al. 1989 Cold Spring Harbor Laboratory Press: ISBN 0-87969-309-6) on a Qiagene DNA preparation robot (Qiagen, Hilden) according to the manufacturer's protocols. Sequencing primers with the following nucleotide sequences were used:

5'-CAGGAAACAGCTATGACC-3'	SEQ ID NO:25
5'-CTAAAGGGAACAAAAGCTG-3'	SEQ ID NO:26

5'-TGTAAAACGACGGCCAGT-3' SEQ ID NO:27

[0146] Sequences were processed and annotated using the software package EST-MAX commercially provided by Bio-Max (Munich, Germany). The program incorporates practically all bioinformatics methods important for functional and structural characterization of protein sequences. For reference the website at *pedant.mips.biochem.mpg.de.* The most important algorithms incorporated in EST-MAX are: FASTA: Very sensitive sequence database searches with estimates of statistical significance; Pearson W.R. (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol. 183:63-98; BLAST: Very sensitive sequence database searches with estimates of statistical significance. Altschul S.F., Gish W., Miller W., Myers E.W., and Lipman D.J. Basic local alignment search tool. Journal of Molecular Biology 215:403-10; PREDATOR: High-accuracy secondary structure prediction from single and multiple sequences. Frishman, D. and Argos, P. (1997) 75% accuracy in protein secondary structure prediction. Proteins, 27:329-335; CLUSTALW: Multiple sequence alignment. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22:4673-4680; TMAP: Transmembrane region prediction from multiply aligned sequences. Persson, B. and Argos, P. (1994) Prediction of transmembrane segments in proteins utilizing multiple sequence alignments. J. Mol. Biol. 237:182-192; ALOM2: Transmembrane region prediction from single sequences. Klein, P., Kanehisa, M., and DeLisi, C. Prediction of protein function from sequence properties: A discriminate analysis of a database. Biochim. Biophys. Acta 787:221-226 (1984). Version 2 by Dr. K. Nakai; PROSEARCH: Detection of PROSITE protein sequence patterns. Kolakowski L.F. Jr., Leunissen J.A.M., Smith J.E. (1992) ProSearch: fast searching of protein sequences with regular expression patterns related to protein structure and function. Biotechniques 13, 919-921; BLIMPS: Similarity searches against a database of ungapped blocks. J.C. Wallace and Henikoff S., (1992); PATMAT: A searching and extraction program for sequence, pattern and block queries and databases, CABIOS 8:249-254. Written by Bill Alford.

Example 5

Identification of Physcomitrella patens ORFs corresponding to APS-2, ZF-2, ZF-3, ZF-4, ZF-5, MYB-1, CABF-3 and SFL-1

[0147] The *Physcomitrella patens* partial cDNAs (ESTs) shown in Table 1 below were identified in the *Physcomitrella patens* EST sequencing program using the program EST-MAX through BLAST analysis. The Sequence Identification Numbers corresponding to these ESTs are as follows: APS-2 (SEQ ID NO:1), ZF-2 (SEQ ID NO:2), ZF-3 (SEQ ID NO:3), ZF-4 (SEQ ID NO:4), ZF-5 (SEQ ID NO:5), MYB-1 (SEQ ID NO:6), CABF-3 (SEQ ID NO:7) and SFL-1 (SEQ ID NO:8).

Table 1

Name	Functional categories	Function	Sequence code	ORF position
PpAPS-2	CBF/Transcription factor	AP2 domain containing protein RAP2.11	c_pp001007077f	592-92
PpZF-2	Transcription factor	zinc finger protein	c_pp004033187r	1688-765
PpZF-3	Transcription factor	BRCA1-associated RING domain protein	c_pp004042321r	1-500
PpZF-4	Transcription factor	zinc finger protein ZNF216	c_pp004059097r	701-1216
PpZF-5	Transcription factor	transcription factor-like protein	c_pp004046041r	1-675
PpMYB-1	Transcription factor	transcription factor	s_pp002016030r	2-505
PpCABF-3	Transcription factor	transcription factor, CCAAT-binding, chain A	c_pp004040113r	221-535
PpSFL-1	Transcription factor	transcription initiation factor sigma A	s_pp001105041r	598-158

Table 2

[0148] Degree of amino acid identity and similarity of PpCBF-3 and other homologous proteins (Pairwise comparison program was used: gap penalty: 10; gap extension penalty: 0.1; score matrix: blosum62)

Swiss-Prot #	O23310	P25209	Q9LFI3	O23633	Q9ZQC3
Protein name	Ccaat-binding transcription factor subunit a	Ccaat-binding transcription factor subunit a	Transcription factor nf-y, ccaat-binding-like protein	Transcriptio n factor	Putative ccaat-binding transcriptio n factor
Species	Arabidopsis thaliana (Mouse-ear cress)	Zea mays (Maize)	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)
Identity %	53%	49%	42%	43%	62%
Similarity %	58%	58%	53%	51%	66%

Table 3

[0149] Degree of amino acid identity and similarity of PpZF-2 and other homologous proteins (Pairwise comparison program was used: gap penalty: 10; gap extension penalty: 0.1; score matrix: blosum62)

Swiss-Prot #	O24008	Q9LUR1	Q9XF63	Q9XF64	Q9LZJ6
Protein name	Zinc finger protein	Ring zinc finger protein-like	Ring-h2 zinc finger protein (atl3)	Ring-h2 zinc finger protein atl5	Ring-h2 zinc finger protein atl5
Species	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)
Identity %	27%	26%	25%	20%	19%
Similarity %	35%	35%	34%	28%	28%

Table 4

[0150] Degree of amino acid identity and similarity of PpZF-3 and other homologous proteins (Pairwise comparison program was used: gap penalty: 10; gap extension penalty: 0.1; score matrix: blosum62)

Swiss-Prot #	Q9SMX5	O04097	Q9UQR3	Q9XZQ1	Q9XZQ2
Protein name	Gcn4-complementing protein (gcp1)	Brca1-associated ring domain proteinisolog	Centaurin beta2	Centaurin beta 1a	Centaurin beta 1b
Species	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)	Homo sapiens (Human)	Caenorhabditis elegans	Caenorhabditis elegans
Identity %	41%	37%	24%	21%	22%
Similarity %	54%	49%	32%	31%	34%

Table 5

[0151] Degree of amino acid identity and similarity of PpZF-4 and other homologous proteins (Pairwise comparison program was used: gap penalty: 10; gap extension penalty: 0.1; score matrix: blosum62)

Swiss-Prot #	Q9LXI5	O88878	O76080	Q9ZNU9	O96038
Protein name	Zinc finger-like protein	Zinc finger protein znf216	Zinc finger protein 216	Putative zinc finger protein	Pem-6
Species	Arabidopsis thaliana (Mouse-ear cress)	Mus musculus (Mouse)	Homo sapiens (Human)	Arabidopsis thaliana (Mouse-ear cress)	Ciona savignyi
Identity %	39%	34%	34%	35%	32%
Similarity %	53%	45%	45%	50%	49%

Table 6

[0152] Degree of amino acid identity and similarity of PpZF-5 and other homologous proteins (Pairwise comparison program was used: gap penalty: 10; gap extension penalty: 0.1; score matrix: blosum62)

Swiss-Prot #	Q9SZW1	Q9ZTR9	Q9SYQ6	Q9ZTX9	O23661
Protein name	Transcription factor-like protein	Auxin response factor 8	Auxin response factor 7	Auxin response factor 4	Ettin protein
Species	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)
Identity %	39%	23%	25%	25%	25%
Similarity %	50%	32%	33%	32%	35%

Table 7

[0153] Degree of amino acid identity and similarity of PpAPS-2 and other homologous proteins (Pairwise comparison program was used: gap penalty: 10; gap extension penalty: 0.1; score matrix: blosum62)

Swiss-Prot #	Q9SJR0	O22174	O04682	Q9SW63	Q9SGJ6
Protein name	Putative ap2 domain transcription factor	Putative ap2 domain containing protein	Pathogenesis-related genes transcriptional activator pti6	Tiny-like protein	Transcription factor dreb1a
Species	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)	Lycopersicon esculentum (Tomato)	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)
Identity %	18%	19%	15%	15%	16%
Similarity %	23%	29%	20%	25%	24%

Table 8

[0154] Degree of amino acid identity and similarity of PpSFL-1 and other homologous proteins (Pairwise comparison program was used: gap penalty: 10; gap extension penalty: 0.1; score matrix: blosum62)

Swiss-Prot #	Q59965	Q9L4T2	O22455	O22056	Q9MTH3
Protein name	Rna polymerase sigma factor	Rna polymerase sigma factor	Rna polymerase sigma factor	Rna polymerase sigma factor	Rna polymerase sigma factor
Species	Synechococcus sp.	Nostoc punctiforme	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)	Sinapis alba (White mustard)
Identity %	49%	49%	32%	42%	30%
Similarity %	62%	61%	44%	59%	42%

Table 9

[0155] Degree of amino acid identity and similarity of PpMYB-1 and other homologous proteins (Pairwise comparison program was used: gap penalty: 10; gap extension penalty: 0.1; score matrix: blosum62).

Swiss-Prot #	Q9LLM9	Q9ZTD9	Q9SEZ4	Q9ZTD7	Q9MBG3
Protein name	Myb-like protein	Putative transcription factor	Putative myb family transcription factor	Putative transcription factor	Myb transcription factor-like protein
Species	Oryza sativa (Rice)	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)	Arabidopsis thaliana (Mouse-ear cress)
Identity %	37%	37%	32%	36%	29%
Similarity %	47%	44%	38%	44%	37%

Example 6

Cloning of the full-length Physcomitrella patens cDNA encoding for APS-2, ZF-2, ZF-3, ZF-4, ZF-5, MYB-1, CABF-3 and SFL-1

[0156] Full-length clones corresponding to CABF-3 (SEQ ID NO:15) and APS-2 (SEQ ID NO:9) were obtained by performing polymerase chain reaction (PCR) with gene-specific primers (see Table 10) and the original EST as the template since they were full-length. The conditions for the reaction are described below under "Full-length Amplification."

[0157] To isolate the clones encoding for PpZF-2, PpZF-3, PpZF-4, PpZF-5 PpAPS-1, PpSFL-1 and PpMYB-1 from *Physcomitrella patens*, cDNA libraries were created with SMART RACE cDNA Amplification kit (Clontech Laboratories) following the manufacturer's instructions. Total RNA isolated as described in Example 3 was used as the template. The cultures were treated prior to RNA isolation as follows: Salt Stress: 2, 6, 12, 24, 48 hours with 1-M NaCl-supplemented medium; Cold Stress: 4°C for the same time points as for salt; Drought Stress: cultures were incubated on dry filter paper for the same time points above. RNA was then pulled and used for isolation.

5' RACE Protocol

[0158] The EST sequences PpZF-2 (SEQ ID NO:2), PpZF-3 (SEQ ID NO:3), PpZF-4 (SEQ ID NO:4), PpZF-5 (SEQ ID NO:5), PpMYB-1 (SEQ ID NO:6) and PpSFL-1 (SEQ ID NO:8) identified from the database search as described in Example 5 were used to design oligos for RACE (see Table 1). The extended sequences for these genes were obtained by performing Rapid Amplification of cDNA Ends polymerase chain reaction (RACE PCR) using the Advantage 2 PCR kit (Clontech Laboratories) and the SMART RACE cDNA amplification kit (Clontech Laboratories) using a Biometra T3 Thermocycler following the manufacturer's instructions.

[0159] The sequences obtained from the RACE reactions contained the 5' end of the full-length coding regions of for PpZF-2, PpZF-3, PpZF-4, PpZF-5 PpAPS-1, PpSFL-1 and PpMYB-1 and were used to design oligos for full-length cloning of the respective genes (see below under "Full-length Amplification).

Full-length Amplification

[0160] Full-length clones corresponding to PpCABF-3 (SEQ ID NO:15) and PpAPS-2 (SEQ ID NO:9) were obtained by performing polymerase chain reaction (PCR) with gene-specific primers (see Table 10) and the original EST as the template. The conditions for the reaction were standard conditions with PWO DNA polymerase (Roche). PCR was performed according to standard conditions and to manufacture's protocols (Sambrook et al. 1989. Molecular Cloning, A Laboratory Manual. 2nd Edition. Cold Spring Harbor Laboratory Press. Cold Spring Harbor,NY, Biometra T3 Thermocycler). The parameters for the reaction were: five minutes at 94°C followed by five cycles of one minute at 94°C, one minute at 50°C and 1.5 minutes at 72°C. This was followed by twenty five cycles of one minute at 94°C, one minute at 65°C and 1.5 minutes at 72°C.

[0161] Full-length clones for PpZF-2 (SEQ ID NO:10), PpZF-3 (SEQ ID NO:11), PpZF-4 (SEQ ID NO:12), PpZF-5 (SEQ ID NO:13), PpMYB-1 (SEQ ID NO:14) and PpSFL-1 (SEQ ID NO:16) and were isolated by repeating the RACE method but using the gene-specific primers as given in Table 10.

[0162] The amplified fragments were extracted from agarose gel with a QIAquick Gel Extraction Kit (Qiagen) and ligated into the TOPO pCR 2.1 vector (Invitrogen) following manufacture's instructions. Recombinant vectors were transformed into Top10 cells (Invitrogen) using standard conditions (Sambrook et al. 1989. Molecular Cloning, A Laboratory Manual. 2nd Edition. Cold Spring Harbor Laboratory Press. Cold Spring Harbor,NY). Transformed cells were selected for on LB agar containing 100 μg/ml

carbenicillin, 0.8mg X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) and 0.8mg IPTG (isopropylthio-β-D-galactoside) grown overnight at 37°C. White colonies were selected and used to inoculate 3ml of liquid LB containing 100 μg/ml ampicillin and grown overnight at 37°C. Plasmid DNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen) following manufacture's instructions. Analyses of subsequent clones and restriction mapping was performed according to standard molecular biology techniques (Sambrook et al. 1989. Molecular Cloning, A Laboratory Manual. 2nd Edition. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY).

Table 10

Gene	**Sites in the final product**	**Isolation Method**	**Primers Race**	**Primer Full-length PCR**
PpCABF-3	XmaI/ SacI	PCR of original EST clone	N/A	RC405 (SEQ ID NO:28) ATCCCGGGCAGCGAG CACACAGCTAGCAAC TCTT RC406 (SEQ ID NO:29) GCGAGCTCACTCCCT CACGCGGTTGACAAT CT
PpZF-2	XmaI/ SacI	5' RACE and RT-PCR for Full-length clone	RC189 (SEQ ID NO:30) TGGCGGCCTC GGTCTTCTTC TCAGT	RC606 (SEQ ID NO:31) ATCCCGGGAGGAAGC TGTCAGGGAAGAGAT GGA RC607 (SEQ ID NO:32) GCGAGCTCTGGCCGT AAAATCAGTTGTGGC GCTT

PpZF-3	XmaI/ EcoRV	5' RACE and RT-PCR for Full-length clone	RC188 (SEQ ID NO:33) CAGCGAAGCC CAATCGGGAT CAGCA	RC604 (SEQ ID NO:34) ATCCCGGGAGGAGG ACTTGCGGAATGCAA ATC RC605 (SEQ ID NO:35) GCGATATCCACCTGC TTCCACTCTCTACTTA TG

PpZF-4	XmaI/ SacI	5' RACE and RT-PCR for Full-length clone	RC185 (SEQ ID NO:36) GACACCCGATTGAGCCGGCAAGACG	RC564 (SEQ ID NO:37) ATCCCGGGCACCAGTCCCGCTTAGTGTGTGTGT RC565 (SEQ ID NO:38) GCGAGCTCTTGATGCGACTCGCTCTCTCGAT
PpZF-5	XmaI/ SacI	5' RACE and RT-PCR for Full-length clone	RC187 (SEQ ID NO:39) CGGCGAGTGCAGCAGCTTCTAGAACG	RC612 (SEQ ID NO:40) ATCCCGGGTATCGATCTGGAGCCCGTTGCAA RC613 (SEQ ID NO:41) GCGAGCTCCTCCAAAGGACTTTGAAATATAGC
PpAPS-2	EcoRV/ SacI	PCR of original EST clone	N/A	RC395(SEQ ID NO:42) GATATCGGAAGAAGAATCCAAGGGAATGCGGTT RC396(SEQ ID NO:43) GCGAGCTCTATGCTTCCGTGGGAGGAGCTTCAC

PpSFL-1	XmaI/ SacI	5' RACE and RT-PCR for Full-length clone	RC172 (SEQ ID NO:44) CCGGCTGGGT TGCCTCAGCT TGCGCA RC538 (SEQ ID NO:45) CGCTCCATCG AACCTGGTGC CTTTGC	RC884 (SEQ ID NO:46) ATCCCGGGCTCGGAA GGACTGTGCATTGTC GA RC885 (SEQ ID NO:47) GCGAGCTCGCAGCAG AAGAAATCCACTTCT GGT
PpMYB-1	SmaI/ SmaI	5' RACE and RT-PCR for Full-length clone	RC170 (SEQ ID NO:48) GGGTGCCGGT TGATGCGAGG GTCCAG	RC701 (SEQ ID NO:49) ATCCCGGGCTGTTGT GTACAGTCTGTGGA RC702 (SEQ ID NO:50) ATCCCGGGCTCACGG AGTAAAGGCCGTACC TT

Example 7

Engineering stress-tolerant Arabidopsis plants by over-expressing the genes APS-2, ZF-2, ZF-3, ZF-4, ZF-5, MYB-1, CABF-3 and SFL-1

Binary vector construction:

[0163] The plasmid construct pACGH101 was digested with PstI (Roche) and FseI (NEB) according to manufacturers' instructions. The fragment was purified by agarose gel and extracted via the Qiaex II DNA Extraction kit (Qiagen). This resulted in a vector fragment with the Arabidopsis Actin2 promoter with internal intron and the OCS3 terminator. Primers for PCR amplification of the NPTII gene were designed as follows:

5'NPT-Pst:

GCG-CTG-CAG-ATT-TCA-TTT-GGA-GAG-GAC-ACG (SEQ ID NO:51)

3'NPT-Fse:

CGC-GGC-CGG-CCT-CAG-AAG-AAC-TCG-TCA-AGA-AGG-CG (SEQ ID NO:52).

[0164] The 0.9 kilobase NPTII gene was amplified via PCR from pCambia 2301 plasmid DNA (94°C for 60 seconds, {94°C for 60 seconds, 61°C (-0.1°C per cycle) for 60 seconds, 72°C for 2 minutes) x 25 cycles, 72°C for 10 minutes on Biometra T-Gradient machine), and purified via the Qiaquick PCR Extraction kit (Qiagen) as per manufacturer's instructions. The PCR DNA was then subcloned into the pCR-BluntII TOPO vector (Invitrogen) pursuant to the manufacturer's instructions (NPT-Topo construct). These ligations were transformed into Top10 cells (Invitrogen) and grown on LB plates with 50 μg/ml kanamycin sulfate overnight at 37°C. Colonies were then used to inoculate 2ml LB media with 50 μg/ml kanamycin sulfate and grown overnight at 37°C. Plasmid DNA was recovered using the Qiaprep Spin Miniprep kit (Qiagen) and sequenced in both the 5' and 3' directions using standard conditions. Subsequent analysis of the sequence data using VectorNTI software revealed no PCR errors present in the NPTII gene sequence.

[0165] The NPT-Topo construct was then digested with PstI (Roche) and FseI (NEB) according to manufacturers' instructions. The 0.9 kilobase fragment was purified on agarose gel and extracted by Qiaex II DNA Extraction kit (Qiagen). The Pst/Fse insert fragment from NPT-Topo and the Pst/Fse vector fragment from pACGH101 were then ligated together using T4 DNA Ligase (Roche) following manufacturer's instructions. The ligation was then transformed into Top10 cells (Invitrogen) under standard conditions, creating pBPSsc019 construct. Colonies were selected on LB plates with 50 μg/ml kanamycin sulfate and grown overnight at 37°C. These colonies were then used to inoculate 2ml LB media with 50 μg/ml kanamycin sulfate and grown overnight at 37°C. Plasmid DNA was recovered using the Qiaprep Spin Miniprep kit (Qiagen) following the manufacturer's instructions.

[0166] The pBPSSC019 construct was digested with KpnI and BsaI (Roche) according to manufacturer's instructions. The fragment was purified via agarose gel and then extracted via the Qiaex II DNA Extraction kit (Qiagen) as per its instructions, resulting in a 3 kilobase Act-NPT cassette, which included the Arabidopsis Actin2 promoter with internal intron, the NPTII gene and the OCS3 terminator.

[0167] The pBPSJH001 vector was digested with SpeI and ApaI (Roche) and blunt-end filled with Klenow enzyme and 0.1mM dNTPs (Roche) according to manufacture's instructions. This produced a 10.1 kilobase vector fragment minus the Gentamycin cassette, which was recircularized by self-ligating with T4 DNA Ligase (Roche), and transformed into

Top10 cells (Invitrogen) via standard conditions. Transformed cells were selected for on LB agar containing 50μg/ml kanamycin sulfate and grown overnight at 37°C. Colonies were then used to inoculate 2ml of liquid LB containing 50μg/ml kanamycin sulfate and grown overnight at 37°C. Plasmid DNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen) following manufacture's instructions. The recircularized plasmid was then digested with KpnI (Roche) and extracted from agarose gel via the Qiaex II DNA Extraction kit (Qiagen) as per manufacturer's instructions.

[0168] The Act-NPT Kpn-cut insert and the Kpn-cut pBPSJH001 recircularized vector were then ligated together using T4 DNA Ligase (Roche) and transformed into Top10 cells (Invitrogen) as per manufacturers' instructions. The resulting construct, pBPSsc022, now contained the Super Promoter, the GUS gene, the NOS terminator, and the Act-NPT cassette. Transformed cells were selected for on LB agar containing 50μg/ml kanamycin sulfate and grown overnight at 37°C. Colonies were then used to inoculate 2ml of liquid LB containing 50μg/ml kanamycin sulfate and grown overnight at 37°C. Plasmid DNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen) following manufacturer's instructions. After confirmation of ligation success via restriction digests, pBPSsc022 plasmid DNA was further propagated and recovered using the Plasmid Midiprep Kit (Qiagen) following the manufacturer's instructions.

Subcloning of APS-2, ZF-2, ZF-3, ZF-4, ZF-5, MYB-1, CABF-3 and SFL-1 into the binary vector

[0169] The fragments containing the different *Physcomitrella patens* transcription factors were subcloned from the recombinant PCR2.1 TOPO vectors by double digestion with restriction enzymes (see Table 11) according to manufacturer's instructions. The subsequence fragment was excised from agarose gel with a QIAquick Gel Extraction Kit (QIAgen) according to manufacture's instructions and ligated into the binary vectors pBPSSC022, cleaved with XmaI and Ecl136II and dephosphorylated prior to ligation. The resulting recombinant pBPSSC022 contained the corresponding transcription factor in the sense orientation under the constitutive super promoter.

Table 11

[0170] Listed are the names of the various constructs of the *Physcomitrella patens* transcription factors used for plant transformation

Gene	Enzymes used to generate gene fragment	Enzymes used to restrict pBPSJH001	Binary Vector Construct
PpCABF-3	XmaI/SacI	XmaI/SacI	pBPSLVM185
PpZF-2	XmaI/SacI	XmaI/SacI	pBPSSY008
PpZF-3	XmaI/ EcoRV	XmaI/Ecl136	pBPSSY017
PpZF-4	XmaI/SacI	XmaI/SacI	pBPSLVM163
PpZF-5	XmaI/SacI	XmaI/SacI	pBPSERG006
PpAPS-2	EcoRV/ SacI	SmaI/SacI	pBPSLVM161
PpSFL-1	XmaI/SacI	XmaI/SacI	pBPSERG001
PpMYB-1	SmaI/ SmaI	SmaI/Ecl136	pBPSERG020

Agrobacterium Transformation

[0171] The recombinant vectors were transformed into *Agrobacterium tumefaciens* C58C1 and PMP90 according to standard conditions (Hoefgen and Willmitzer, 1990).

Plant Transformation

[0172] *Arabidopsis thaliana* ecotype C24 were grown and transformed according to standard conditions (Bechtold 1993, Acad. Sci. Paris. 316:1194-1199; Bent et al. 1994, Science 265:1856-1860).

Screening of Transformed Plants

[0173] T1 seeds were sterilized according to standard protocols (Xiong et al. 1999, Plant Molecular Biology Reporter 17: 159-170). Seeds were plated on ½ Murashige and Skoog media (MS) (Sigma-Aldrich) pH 5.7 with KOH, 0.6% agar and supplemented with 1% sucrose, 0.5 g/L 2-[N-Morpholino]ethansulfonic acid (MES) (Sigma-Aldrich), 50 µg/ml kanamycin (Sigma-Aldrich), 500 µg/ml carbenicillan (Sigma-Aldrich) and 2 µg/ml benomyl (Sigma-Aldrich). Seeds on plates were vernalized for four days at 4°C. The seeds were germinated in a climatic chamber at an air temperature of 22°C and light intensity of 40 micromols^{-1}m2 (white light; Philips TL 65W/25 fluorescent tube) and 16 hours light and 8 hours dark day length cycle. Transformed seedlings were selected after 14 days and

transferred to ½ MS media pH 5.7 with KOH 0.6% agar plates supplemented with 0.6% agar, 1% sucrose, 0.5 g/L MES (Sigma-Aldrich), and 2 μg/ml benomyl (Sigma-Aldrich) and allowed to recover for five-seven days.

Drought Tolerance Screening

[0174] T1 seedlings were transferred to dry, sterile filter paper in a petri dish and allowed to desiccate for two hours at 80% RH (relative humidity) in a Percival Growth CU3615, micromols^{-1}m2 (white light; Philips TL 65W/25 fluorescent tube). The RH was then decreased to 60% and the seedlings were desiccated further for eight hours. Seedlings were then removed and placed on ½ MS 0.6% agar plates supplemented with 2μg/ml benomyl (Sigma-Aldrich) and 0.5g/L MES ((Sigma-Aldrich) and scored after five days.

[0175] Under drought stress conditions, PpCABF-3 over-expressing Arabidopsis thaliana plants showed an 70% (39 survivors from 56 stressed plants) survival rate to the stress screening; PpZF-2, 98% (39 survivors from 40 stressed plants); PpZF-3, 94% (59 survivors from 63 stressed plants); PpZF-4, 94% (16 survivors from 17 stressed plants); PpZF-5, 80% (8 survivors from 10 stressed plants); PpAPS-2 65% (13 survivors from 20 stressed plants); and PpMYB-1 80% (8 survivors from 10 stressed plants); whereas the untransformed control a 28% (16 survivors from 57 stressed plants) survival rate. It is noteworthy that the analyses of these transgenic lines were performed with T1 plants, and therefore, the results will be better when a homozygous, strong expresser is found.

Table 12

[0176] Summary of the drought stress tests

Gene Name	Drought Stress Test		
	Number of survivors	Total number of plants	Percentage of survivors
PpCABF-3	39	56	70%
PpZF-2	39	40	98%
PpZF-3	59	63	94%
PpZF-4	16	17	94%
PpZF-5	8	10	80%
PpAPS-2	13	20	65
PpMYB-1	8	10	80%
Control	16	57	28%

Freezing Tolerance Screening

[0177] Seedlings were moved to petri dishes containing ½ MS 0.6% agar supplemented with 2% sucrose and 2 μg/ml benomyl. After four days, the seedlings were incubated at 4°C for 1 hour and then covered with shaved ice. The seedlings were then placed in an Environmental Specialist ES2000 Environmental Chamber and incubated for 3.5 hours beginning at –1.0°C decreasing –1°C hour. The seedlings were then incubated at –5.0°C for 24 hours and then allowed to thaw at 5°C for 12 hours. The water was poured off and the seedlings were scored after 5 days.

[0178] Under freezing stress conditions, PpCABF-3 over-expressing Arabidopsis thaliana plants showed an 98% (41 survivors from 42 stressed plants) survival rate to the stress screening; PpZF-2, 86% (19 survivors from 22 stressed plants); and PpZF-3, 74% (14 survivors from 19 stressed plants); whereas the untransformed control a 28% (16 survivors from 57 stressed plants) survival rate. It is noteworthy that the analyses of these transgenic lines were performed with T1 plants, and therefore, the results will be better when a homozygous, strong expresser is found.

Table 13

[0179] Summary of the freezing stress tests

Gene Name	Freezing Stress Test		
	Number of survivors	Total number of plants	Percentage of survivors
PpCABF-3	41	42	98%
PpZF-2	19	22	86%
PpZF-3	14	19	74%
Control	1	48	2%

Salt Tolerance Screening

[0180] Seedlings were transferred to filter paper soaked in ½ MS and placed on ½ MS 0.6% agar supplemented with 2 μg/ml benomyl the night before the salt tolerance screening. For the salt tolerance screening, the filter paper with the seedlings was moved to stacks of sterile filter paper, soaked in 50 mM NaCl, in a petri dish. After two hours, the filter paper with the seedlings was moved to stacks of sterile filter paper, soaked with 200 mM NaCl, in a petri dish. After two hours, the filter paper with the seedlings was moved to stacks of sterile filter paper, soaked in 600mM NaCl, in a petri dish. After 10 hours, the seedlings were moved to petri dishes containing ½ MS 0.6% agar supplemented with 2 μg/ml benomyl. The seedlings were scored after 5 days.

[0181] The transgenic plants are screened for their improved salt tolerance demonstrating that transgene expression confers salt tolerance.

Example 8

Detection of the APS-2, ZF-2, ZF-3, ZF-4, ZF-5, MYB-1, CABF-3 and SFL-1 transgenes in the transgenic Arabidopsis lines

[0182] One leaf from a wild type and a transgenic *Arabidopsis* plant was homogenized in 250 μl Hexadecyltrimethyl ammonium bromide (CTAB) buffer (2% CTAB, 1.4 M NaCl, 8mM EDTA and 20mM Tris pH 8.0) and 1 μl β-mercaptoethanol. The samples were incubated at 60-65°C for 30 minutes and 250 μl of Chloroform was then added to each sample. The samples were vortexed for 3 minutes and centrifuged for 5 minutes at 18,000 x *g*. The supernatant was taken from each sample and 150 μl isopropanol was added. The

samples were incubated at room temperature for 15 minutes, and centrifuged for 10 minutes at 18,000 x *g*. Each pellet was washed with 70% ethanol, dried, and resuspended in 20 µl TE. 4 µl of above suspension was used in a 20 µl PCR reaction using *Taq* DNA polymerase (Roche Molecular Biochemicals) according to the manufacturer's instructions. Binary vector plasmid with each gene cloned in was used as positive control, and the wild type C24 genomic DNA was used as negative control in the PCR reactions. 10 µl PCR reaction was analyzed on 0.8% agarose/ethidium bromide gel. The PCR program used was as follows: 30 cycles of 1 minute at 94°C, 1 minute at 62°C and 4 minutes at 70°C, followed by 10 minutes at 72°C. The following primer was used as 5' primer: Bfwd: 5'GCTGACACGCCAAGCCTCGCTAGTC3'. (SEQ ID NO:53) The gene-specific primers and the size of the amplified bands (Gene Product Size) are listed below.

[0183] PpCABF-3

Primer: RC406: GCGAGCTCACTCCCTCACGCGGTTGACAATCT

Gene Product Size: 800 bp (SEQ ID NO:54)

[0184] PpZF-2

Primer: RC607: GCGAGCTCTGGCCGTAAAATCAGTTGTGGCGCTT

Gene Product Size: 1800 bp (SEQ ID NO:55)

[0185] PpZF-3

Primer: RC605: GCGATATCCACCTGCTTCCACTCTCTACTTATG

Gene Product Size: 2000 bp (SEQ ID NO:56)

[0186] PpZF-4

Primer: RC565: GCGAGCTCTTGATGCGACTCGCTCTCTCGAT

Gene Product Size: 800 bp (SEQ ID NO:57)

[0187] PpZF-5

Primer: RC613: GCGAGCTCCTCCAAAGGACTTTGAAATATAGC

Gene Product Size: 2700 bp (SEQ ID NO:58)

[0188] PpAPS-2

Primer: RC396: GCGAGCTCTATGCTTCCGTGGGAGGAGCTTCAC
Gene Product Size: 1000 bp (SEQ ID NO:59)

[0189] PpSFL-1
Primer: RC885: GCGAGCTCGCAGCAGAAGAAATCCACTTCTGGT
Gene Product Size: 1700 bp (SEQ ID NO:60)

[0190] PpMYB-1
Primer: RC702: ATCCCGGGCTCACGGAGTAAAGGCCGTACCTT
Gene Product Size: 2400 bp (SEQ ID NO:61)

[0191] The transgenes were successfully amplified from the T1 transgenic lines, but not from the wild type C24. This result indicates that the T1 transgenic plants contain at least one copy of the transgenes. There was no indication of existence of either identical or very similar genes in the untransformed Arabidopsis thaliana control which could be amplified by this method.

Example 9

Detection of the APS-2, ZF-2, ZF-3, ZF-4, ZF-5, MYB-1, CABF-3 and SFL-1 transgene mRNA in transgenic Arabidopsis lines

[0192] Transgene expression was detected using RT-PCR. Total RNA was isolated from stress-treated plants using a procedure adapted from (Verwoerd et al., 1989 NAR 17:2362). Leaf samples (50-100 mg) were collected and ground to a fine powder in liquid nitrogen. Ground tissue was resuspended in 500 µl of a 80°C, 1:1 mixture, of phenol to extraction buffer (100mM LiCl, 100 mM Tris pH8, 10 mM EDTA, 1% SDS), followed by brief vortexing to mix. After the addition of 250 µl of chloroform, each sample was vortexed briefly. Samples were then centrifuged for 5 minutes at 12,000 x g. The upper aqueous phase was removed to a fresh eppendorf tube. RNA was precipitated by adding $1/10^{th}$ volume 3M sodium acetate and 2 volumes 95% ethanol. Samples were mixed by inversion and placed on ice for 30 minutes. RNA was pelleted by centrifugation at 12,000 x g for 10 minutes. The supernatant was removed and pellets briefly air-dried. RNA sample pellets were resuspended in 10 µl DEPC treated water.

[0193] To remove contaminating DNA from the samples, each was treated with RNase-free DNase (Roche) according to the manufacturer's recommendations. cDNA was synthesized from total RNA using the 1st Strand cDNA synthesis kit (Boehringer Mannheim) following manufacturer's recommendations. PCR amplification of a gene-specific fragment from the synthesized cDNA was performed using *Taq* DNA polymerase (Roche) and gene-specific primers (see Table 4 for primers) in the following reaction: 1X PCR buffer, 1.5mM $MgCl_2$, 0.2 μM each primer, 0.2μM dNTPs, 1 unit polymerase, 5μl cDNA from synthesis reaction. Amplification was performed under the following conditions: Denaturation, 95°C, 1 minute; annealing, 62°C, 30 seconds; extension, 72°C, 1 minute, 35 cycles; extension, 72°C, 5 minutes; hold, 4°C, forever. PCR products were run on a 1% agarose gel, stained with ethidium bromide, and visualized under UV light using the Quantity-One gel documentation system (Bio-Rad).Expression of the transgenes was detected in the T1 transgenic line.

[0194] These results indicated that the transgenes are expressed in the transgenic lines and strongly suggested that their gene product improved plant stress tolerance in the transgenic lines. In agreement with the previous statement, no expression of identical or very similar endogenous genes could be detected by this method. These results are in agreement with the data from Example 7.

Table 14

[0195] Primers used for the amplification of respective transgene mRNA in PCR using RNA isolated from transgenic Arabidopsis thaliana plants as template

Gene	5' primer	3' primer
PpCABF-2	RC405: (SEQ ID NO:62) ATCCCGGGCAGCGAGC ACACAGCTAGCAACTC TT	RC406: (SEQ ID NO:63) GCGAGCTCACTCCCTC ACGCGGTTGACAATCT
PpZF-2	RC1191: (SEQ ID NO:64) GCCCGTTGTGTCGCAC GAGTGTGGGA	RC1192: (SEQ ID NO:65) GCCGCTGGACCAGACC TCGGAATGT
PpZF-3	RC1203: (SEQ ID NO:66) GAGGCAGTCATGCAAT CGACCCCAA	RC1204: (SEQ ID NO:67) GCGAAGCCCAATCGGG ATCAGCAGCA
PpZF-4	RC564: (SEQ ID NO:68) ATCCCGGGCACCAGTC CCGCTTAGTGTGTGTGT	RC565: (SEQ ID NO:69) GCGAGCTCTTGATGCG ACTCGCTCTCTCGAT
PpZF-5	RC1209: (SEQ ID NO:70) CGCATCGCATCTGGCG AACTTTGTG	RC1210: (SEQ ID NO:71) 3' primer for EST281 at#1368 GC=58% CGTACCACGATTGCTCT AGCGCACGT
PpAPS-1	RC395: (SEQ ID NO:72) GCGATATCGGAAGAAG AATCCAAGGGAATGCG GTT	RC396: (SEQ ID NO:73) GCGAGCTCTATGCTTCC GTGGGAGGAGCTTCAC
PpAPS-	RC405: (SEQ ID NO:74) ATCCCGGGCAGCGAGC ACACAGCTAGCAACTC TT	RC406: (SEQ ID NO:75) GCGAGCTCACTCCCTC ACGCGGTTGACAATCT
PpSFL-1	RC1191: (SEQ ID NO:76) GCCCGTTGTGTCGCAC GAGTGTGGGA	RC1192: (SEQ ID NO:77) GCCGCTGGACCAGACC TCGGAATGT

PpMYB-1	RC1203: (SEQ ID NO:78) GAGGCAGTCATGCAAT CGACCCCAA	RC1204: (SEQ ID NO:79) GCGAAGCCCAATCGGG ATCAGCAGCA

Example 10

Engineering stress-tolerant soybean plants by over-expressing the APS-2, ZF-2, ZF-3, ZF-4, ZF-5, MYB-1, CABF-3 and SFL-1 gene

[0196] The constructs pBPSLVM185, pBPSSY008, pBPSSY017, pBPSLVM163, pBPSERG006, pBPSLVM161, pBPSERG001 and pBPSERG020 were used to transform soybean as described below.

[0197] Seeds of soybean were surface sterilized with 70% ethanol for 4 minutes at room temperature with continuous shaking, followed by 20% (v/v) Clorox supplemented with 0.05% (v/v) Tween for 20 minutes with continuous shaking. Then, the seeds were rinsed 4 times with distilled water and placed on moistened sterile filter paper in a Petri dish at room temperature for 6 to 39 hours. The seed coats were peeled off, and cotyledons are detached from the embryo axis. The embryo axis was examined to make sure that the meristematic region is not damaged. The excised embryo axes were collected in a half-open sterile Petri dish and air-dried to a moisture content less than 20% (fresh weight) in a sealed Petri dish until further use.

[0198] *Agrobacterium tumefaciens* culture was prepared from a single colony in LB solid medium plus appropriate antibiotics (e.g. 100 mg/l streptomycin, 50 mg/l kanamycin) followed by growth of the single colony in liquid LB medium to an optical density at 600 nm of 0.8. Then, the bacteria culture was pelleted at 7000 rpm for 7 minutes at room temperature, and resuspended in MS (Murashige and Skoog, 1962) medium supplemented with 100 μM acetosyringone. Bacteria cultures were incubated in this pre-induction medium for 2 hours at room temperature before use. The axis of soybean zygotic seed embryos at approximately 15% moisture content were imbibed for 2 hours at room temperature with the pre-induced *Agrobacterium* suspension culture. The embryos are removed from the imbibition culture and were transferred to Petri dishes containing solid MS medium supplemented with 2% sucrose and incubated for 2 days, in the dark at room temperature. Alternatively, the embryos were placed on top of moistened (liquid MS medium) sterile filter paper in a Petri dish and incubated under the same conditions described above. After this period, the embryos were transferred to either solid or liquid MS medium supplemented with 500 mg/L carbenicillin or 300mg/L cefotaxime to kill the agrobacteria. The liquid medium was used to moisten the

sterile filter paper. The embryos were incubated during 4 weeks at 25°C, under 150 μmol $m^{-2}sec^{-1}$ and 12 hours photoperiod. Once the seedlings produced roots, they were transferred to sterile metromix soil. The medium of the *in vitro* plants was washed off before transferring the plants to soil. The plants were kept under a plastic cover for 1 week to favor the acclimatization process. Then the plants were transferred to a growth room where they were incubated at 25°C, under 150 μmol $m^{-2}sec^{-1}$ light intensity and 12 hours photoperiod for about 80 days.

[0199] The transgenic plants were then screened for their improved drought, salt and/or cold tolerance according to the screening method described in Example 7 demonstrating that transgene expression confers stress tolerance.

Example 11

Engineering stress-tolerant Rapeseed/Canola plants by over-expressing the APS-2, ZF-2, ZF-3, ZF-4, ZF-5, MYB-1, CABF-3 and SFL-1 genes

[0200] The constructs pBPSLVM185, pBPSSY008, pBPSSY017, pBPSLVM163, pBPSERG006, pBPSLVM161, pBPSERG001 and pBPSERG020 were used to transform rapseed/canola as described below.

[0201] The method of plant transformation described herein is also applicable to *Brassica* and other crops. Seeds of canola are surface sterilized with 70% ethanol for 4 minutes at room temperature with continuous shaking, followed by 20% (v/v) Clorox supplemented with 0.05 % (v/v) Tween for 20 minutes, at room temperature with continuous shaking. Then, the seeds are rinsed 4 times with distilled water and placed on moistened sterile filter paper in a Petri dish at room temperature for 18 hours. Then the seed coats are removed and the seeds are air dried overnight in a half-open sterile Petri dish. During this period, the seeds lose approx. 85% of its water content. The seeds are then stored at room temperature in a sealed Petri dish until further use. DNA constructs and embryo imbibition are as described in Example 10. Samples of the primary transgenic plants (T0) are analyzed by PCR to confirm the presence of T-DNA. These results are confirmed by Southern hybridization in which DNA is electrophoresed on a 1% agarose gel and transferred to a positively charged nylon membrane (Roche Diagnostics). The PCR DIG Probe Synthesis Kit (Roche Diagnostics) is used to prepare a digoxigenin-labelled probe by PCR, and used as recommended by the manufacturer.

[0202] The transgenic plants are then screened for their improved stress tolerance according to the screening method described in Example 7 demonstrating that transgene expression confers drought tolerance.

Example 12

Engineering stress-tolerant corn plants by over-expressing the APS-2, ZF-2, ZF-3, ZF-4, ZF-5, MYB-1, CABF-3 or SFL-1 genes

[0203] The constructs pBPSLVM185, pBPSSY008, pBPSSY017, pBPSLVM163, pBPSERG006, pBPSLVM161, pBPSERG001 and pBPSERG020 were used to transform corn as described below.

[0204] Transformation of maize (*Zea Mays L.*) is performed with the method described by Ishida et al. 1996. Nature Biotch 14745-50. Immature embryos are co-cultivated with *Agrobacterium tumefaciens* that carry "super binary" vectors, and transgenic plants are recovered through organogenesis. This procedure provides a transformation efficiency of between 2.5% and 20%. The transgenic plants are then screened for their improved drought, salt and/or cold tolerance according to the screening method described in Example 7 demonstrating that transgene expression confers stress tolerance.

Example 13

Engineering stress-tolerant wheat plants by over-expressing the APS-2, ZF-2, ZF-3, ZF-4, ZF-5, MYB-1, CABF-3 or SFL-1 genes

[0205] The constructs pBPSLVM185, pBPSSY008, pBPSSY017, pBPSLVM163, pBPSERG006, pBPSLVM161, pBPSERG001, pBPSERG020 were used to transform wheat as described below.

[0206] Transformation of wheat is performed with the method described by Ishida et al. 1996 Nature Biotch. 14745-50. Immature embryos are co-cultivated with *Agrobacterium tumefaciens* that carry "super binary" vectors, and transgenic plants are recovered through organogenesis. This procedure provides a transformation efficiency between 2.5% and 20%. The transgenic plants are then screened for their improved stress tolerance according to the screening method described in Example 7 demonstrating that transgene expression confers drought tolerance.

Example 14

Identification of Homologous and Heterologous Genes

[0207] Gene sequences can be used to identify homologous or heterologous genes from cDNA or genomic libraries. Homologous genes (e. g. full-length cDNA clones) can be isolated via nucleic acid hybridization using for example cDNA libraries. Depending on the abundance of the gene of interest, 100,000 up to 1,000,000 recombinant bacteriophages are plated and transferred to nylon membranes. After denaturation with alkali, DNA is immobilized on the membrane by e. g. UV cross linking. Hybridization is carried out at high stringency conditions. In aqueous solution hybridization and washing is performed at an ionic strength of 1 M NaCl and a temperature of 68°C. Hybridization probes are generated by e. g. radioactive (^{32}P) nick transcription labeling (High Prime, Roche, Mannheim, Germany). Signals are detected by autoradiography.

[0208] Partially homologous or heterologous genes that are related but not identical can be identified in a manner analogous to the above-described procedure using low stringency hybridization and washing conditions. For aqueous hybridization, the ionic strength is normally kept at 1 M NaCl while the temperature is progressively lowered from 68 to 42°C.

[0209] Isolation of gene sequences with homologies (or sequence identity/similarity) only in a distinct domain of (for example 10-20 amino acids) can be carried out by using synthetic radio labeled oligonucleotide probes. Radio labeled oligonucleotides are prepared by phosphorylation of the 5-prime end of two complementary oligonucleotides with T4 polynucleotide kinase. The complementary oligonucleotides are annealed and ligated to form concatemers. The double stranded concatemers are than radiolabeled by, for example, nick transcription. Hybridization is normally performed at low stringency conditions using high oligonucleotide concentrations.

Oligonucleotide hybridization solution:
6 x SSC
0.01 M sodium phosphate
1 mM EDTA (pH 8)
0.5 % SDS
100 µg/ml denatured salmon sperm DNA
0.1 % nonfat dried milk

[0210] During hybridization, temperature is lowered stepwise to 5-10°C below the estimated oligonucleotide Tm or down to room temperature followed by washing steps and autoradiography. Washing is performed with low stringency such as 3 washing steps using 4 x SSC. Further details are described by Sambrook, J. *et al.* (1989), "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons.

Example 15

Identification of Homologous Genes by Screening Expression Libraries with Antibodies

[0211] c-DNA clones can be used to produce recombinant protein for example in E. coli (e. g. Qiagen QIAexpress pQE system). Recombinant proteins are then normally affinity purified via Ni-NTA affinity chromatography (Qiagen). Recombinant proteins are then used to produce specific antibodies for example by using standard techniques for rabbit immunization. Antibodies are affinity purified using a Ni-NTA column saturated with the recombinant antigen as described by Gu et al., 1994 BioTechniques 17:257-262. The antibody can than be used to screen expression cDNA libraries to identify homologous or heterologous genes via an immunological screening (Sambrook, J. *et al.* (1989), "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press or Ausubel, F.M. *et al.* (1994) "Current Protocols in Molecular Biology", John Wiley & Sons).

Example 16

In vivo Mutagenesis

[0212] *In vivo* mutagenesis of microorganisms can be performed by passage of plasmid (or other vector) DNA through *E. coli* or other microorganisms (e.g. *Bacillus* spp. or yeasts such as *Saccharomyces cerevisiae*) which are impaired in their capabilities to maintain the integrity of their genetic information. Typical mutator strains have mutations in the genes for the DNA repair system (e.g., mutHLS, mutD, mutT, etc.; for reference, see Rupp, W.D. (1996) DNA repair mechanisms, in: *Escherichia coli* and *Salmonella*, p. 2277-2294, ASM: Washington.) Such strains are well known to those skilled in the art. The use of such strains is illustrated, for example, in Greener, A. and Callahan, M. (1994) *Strategies* 7: 32-34. Transfer of mutated DNA molecules into plants is preferably done after selection and testing in microorganisms. Transgenic plants are generated according to various examples within the exemplification of this document.

Example 17

In vitro Analysis of the Function of Physcomitrella Genes in Transgenic Organisms

[0213] The determination of activities and kinetic parameters of enzymes is well established in the art. Experiments to determine the activity of any given altered enzyme must be tailored to the specific activity of the wild-type enzyme, which is well within the ability of one skilled in the art. Overviews about enzymes in general, as well as specific details concerning structure, kinetics, principles, methods, applications and examples for the determination of many enzyme activities may be found, for example, in the following references: Dixon, M., and Webb, E.C., (1979) Enzymes. Longmans: London; Fersht, (1985) Enzyme Structure and Mechanism. Freeman: New York; Walsh, (1979) Enzymatic Reaction Mechanisms. Freeman: San Francisco; Price, N.C., Stevens, L. (1982) Fundamentals of Enzymology. Oxford Univ. Press: Oxford; Boyer, P.D., ed. (1983) The Enzymes, 3rd ed. Academic Press: New York; Bisswanger, H., (1994) Enzymkinetik, 2nd ed. VCH: Weinheim (ISBN 3527300325); Bergmeyer, H.U., Bergmeyer, J., Graßl, M., eds. (1983-1986) Methods of Enzymatic Analysis, 3rd ed., vol. I-XII, Verlag Chemie: Weinheim; and Ullmann's Encyclopedia of Industrial Chemistry (1987) vol. A9, Enzymes. VCH: Weinheim, p. 352-363.

[0214] The activity of proteins which bind to DNA can be measured by several well-established methods, such as DNA band-shift assays (also called gel retardation assays). The effect of such proteins on the expression of other molecules can be measured using reporter gene assays (such as that described in Kolmar, H. et al. (1995) *EMBO J.* 14: 3895-3904 and references cited therein). Reporter gene test systems are well known and established for applications in both pro- and eukaryotic cells, using enzymes such as β-galactosidase, green fluorescent protein, and several others.

[0215] The determination of activity of membrane-transport proteins can be performed according to techniques such as those described in Gennis, R.B. Pores, Channels and Transporters, in Biomembranes, Molecular Structure and Function, pp. 85-137, 199-234 and 270-322, Springer: Heidelberg (1989).

Example 18

Purification of the Desired Product from Transformed Organisms

[0216] Recovery of the desired product from plant material (i.e., *Physcomitrella patents* or *Arabidopsis thaliana*), fungi, algae, ciliates, *C. glutamicum* cells, or other bacterial

cells transformed with the nucleic acid sequences described herein, or the supernatant of the above-described cultures can be performed by various methods well known in the art. If the desired product is not secreted from the cells, can be harvested from the culture by low-speed centrifugation, the cells can be lysed by standard techniques, such as mechanical force or sonification. Organs of plants can be separated mechanically from other tissue or organs. Following homogenization cellular debris is removed by centrifugation, and the supernatant fraction containing the soluble proteins is retained for further purification of the desired compound. If the product is secreted from desired cells, then the cells are removed from the culture by low-speed centrifugation, and the supernate fraction is retained for further purification.

[0217] The supernatant fraction from either purification method is subjected to chromatography with a suitable resin, in which the desired molecule is either retained on a chromatography resin while many of the impurities in the sample are not, or where the impurities are retained by the resin while the sample is not. Such chromatography steps may be repeated as necessary, using the same or different chromatography resins. One skilled in the art would be well-versed in the selection of appropriate chromatography resins and in their most efficacious application for a particular molecule to be purified. The purified product may be concentrated by filtration or ultrafiltration, and stored at a temperature at which the stability of the product is maximized.

[0218] There is a wide array of purification methods known to the art and the preceding method of purification is not meant to be limiting. Such purification techniques are described, for example, in Bailey, J.E. & Ollis, D.F. Biochemical Engineering Fundamentals, McGraw-Hill: New York (1986). Additionally, the identity and purity of the isolated compounds may be assessed by techniques standard in the art. These include high-performance liquid chromatography (HPLC), spectroscopic methods, staining methods, thin layer chromatography, NIRS, enzymatic assay, or microbiologically. Such analysis methods are reviewed in: Patek et al., 1994 *Appl. Environ. Microbiol.* 60:133-140; Malakhova et al., 1996 *Biotekhnologiya* 11:27-32; and Schmidt et al., 1998 *Bioprocess Engineer.* 19:67-70. Ulmann's Encyclopedia of Industrial Chemistry, (1996) vol. A27, VCH: Weinheim, p. 89-90, p. 521-540, p. 540-547, p. 559-566, 575-581 and p. 581-587; Michal, G. (1999) Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley and Sons; Fallon, A. et al. (1987) Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology, vol. 17.

CLAIMS

WE CLAIM:

1. A transgenic plant cell transformed by a Transcription Factor Stress-Related Protein (TFSRP) coding nucleic acid, wherein expression of the nucleic acid in the plant cell results in increased tolerance to an environmental stress as compared to a wild type variety of the plant cell and wherein the TFSRP is selected from the group consisting of 1) a AP2 Similar-2 protein (APS-2); 2) a Zinc-Finger Factor-2 protein (ZF-2); 3) a Zinc-Finger Factor-3 protein (ZF-3); 4) a Zinc-Finger Factor-4 protein (ZF-4); 5) a Zinc-Finger Factor-5 protein (ZF-5); 6) a MYB-1 protein (MYB-1); 7) a CAAT-Box Binding Factor protein-3 (CABF-3); 8) a Sigma Factor Like protein (SFL-1); and orthologs thereof.

2. The transgenic plant cell of Claim 1, wherein the TFSRP is selected from the group consisting of 1) APS-2 as defined in SEQ ID NO:17; 2) ZF-2 as defined in SEQ ID NO:18; 3) ZF-3 as defined in SEQ ID NO:19; 4) ZF-4 as defined in SEQ ID NO:20; 5) ZF-5 as defined in SEQ ID NO:21; 6) MYB-1 as defined in SEQ ID NO:22; 7) CABF-3 as defined in SEQ ID NO:23 and 8) SFL-1 as defined in SEQ ID NO:24.

3. The transgenic plant cell of Claim 1, wherein the TFSRP coding nucleic acid is selected from the group consisting of 1) APS-2 as defined in SEQ ID NO:9; 2) ZF-2 as defined in SEQ ID NO:10; 3) ZF-3 as defined in SEQ ID NO:11; 4) ZF-4 as defined in SEQ ID NO:12; 5) ZF-5 as defined in SEQ ID NO:13; 6) MYB-1 as defined in SEQ ID NO:14; 7) CABF-3 as defined in SEQ ID NO:15 and 8) SFL-1 as defined in SEQ ID NO:16.

4. The transgenic plant cell of Claim 1, wherein the TFSRP coding nucleic acid hybridizes under stringent conditions to a sequence of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16.

5. The transgenic plant cell of Claim 1, wherein the environmental stress is selected from the group consisting of salinity, drought and temperature.

6. The transgenic plant cell of Claim 1, wherein the plant is a monocot.

7. The transgenic plant cell of Claim 1, wherein the plant is a dicot.

8. The transgenic plant cell of Claim 1, wherein the plant is selected from the group consisting of maize, wheat, rye, oat, triticale, rice, barley, soybean, peanut, cotton, rapeseed, canola, manihot, pepper, sunflower, tagetes, solanaceous plants, potato, tobacco, eggplant, tomato, Vicia species, pea, alfalfa, coffee, cacao, tea, Salix species, oil palm, coconut, perennial grass and forage crops.

9. A transgenic plant comprising a plant cell according to any of Claims 1-8.

10. A seed produced by a transgenic plant comprising a plant cell according to any of Claims 1-8, wherein the seed is true breeding for an increased tolerance to environmental stress as compared to a wild type variety of the plant cell.

11. An agricultural product produced by the transgenic plant or seed of Claims 9 or 10.

12. An isolated Transcription Factor Stress-Related Protein (TFSRP) selected from the group consisting of: 1) a AP2 Similar-2 protein (APS-2); 2) a Zinc-Finger Factor-2 protein (ZF-2); 3) a Zinc-Finger Factor-3 protein (ZF-3); 4) a Zinc-Finger Factor-4 protein (ZF-4); 5) a Zinc-Finger Factor-5 protein (ZF-5); 6) a MYB-1 protein (MYB-1); 7) a CAAT-Box Binding Factor protein-3 (CABF-3); 8) a Sigma Factor Like protein (SFL-1); and orthologs thereof.

13. The isolated TFSRP of Claim 12, wherein the TFSRP is selected from the group consisting of: 1) APS-2 as defined in SEQ ID NO:17; 2) ZF-2 as defined in SEQ ID NO:18; 3) ZF-3 as defined in SEQ ID NO:19; 4) ZF-4 as defined in SEQ ID NO:20; 5) ZF-5 as defined in SEQ ID NO:21; 6) MYB-1 as defined in SEQ ID NO:22; 7) CABF-3 as defined in SEQ ID NO:23 and 8) SFL-1 as defined in SEQ ID NO:24.

14. An isolated Transcription Factor Stress-Related Protein (TFSRP) coding nucleic acid, wherein the TFSRP coding nucleic acid codes for a TFSRP selected from the group consisting of: 1) a AP2 Similar-2 protein (APS-2); 2) a Zinc-Finger Factor-2 protein (ZF-2); 3) a Zinc-Finger Factor-3 protein (ZF-3); 4) a Zinc-Finger Factor-4 protein (ZF-4); 5) a Zinc-Finger Factor-5 protein (ZF-5); 6) a MYB-1 protein (MYB-1); 7) a CAAT-Box Binding Factor protein-3 (CABF-3); 8) a Sigma Factor Like protein (SFL-1); and orthologs thereof.

15. The TFSRP coding nucleic acid of Claim 14, wherein the TFSRP coding nucleic acid selected from the group consisting of 1) APS-2 as defined in SEQ ID NO:9; 2) ZF-2

as defined in SEQ ID NO:10; 3) ZF-3 as defined in SEQ ID NO:11; 4) ZF-4 as defined in SEQ ID NO:12; 5) ZF-5 as defined in SEQ ID NO:13; 6) MYB-1 as defined in SEQ ID NO:14; 7) CABF-3 as defined in SEQ ID NO:15 and 8) SFL-1 as defined in SEQ ID NO:16.

16. An isolated recombinant expression vector comprising a nucleic acid of Claims 14 or 15, wherein expression of the vector in a host cell results in increased tolerance to environmental stress as compared to a wild type variety of the host cell.

17. A method of producing a transgenic plant containing a Transcription Factor Stress-Related Protein (TFSRP) coding nucleic acid, wherein expression of the nucleic acid in the plant results in increased tolerance to environmental stress as compared to a wild type variety of the plant, comprising, transforming a plant cell with an expression vector comprising the nucleic acid, generating from the plant cell a transgenic plant with an increased tolerance to environmental stress as compared to a wild type variety of the plant, wherein the TFSRP is selected from the group consisting of 1) a AP2 Similar-2 protein (APS-2); 2) a Zinc-Finger Factor-2 protein (ZF-2); 3) a Zinc-Finger Factor-3 protein (ZF-3); 4) a Zinc-Finger Factor-4 protein (ZF-4); 5) a Zinc-Finger Factor-5 protein (ZF-5); 6) a MYB-1 protein (MYB-1); 7) a CAAT-Box Binding Factor protein-3 (CABF-3); 8) a Sigma Factor Like protein (SFL-1); and orthologs thereof.

18. The method of Claim 17, wherein the TFSRP is selected from the group consisting of 1) APS-2 as defined in SEQ ID NO:17; 2) ZF-2 as defined in SEQ ID NO:18; 3) ZF-3 as defined in SEQ ID NO:19; 4) ZF-4 as defined in SEQ ID NO:20; 5) ZF-5 as defined in SEQ ID NO:21; 6) MYB-1 as defined in SEQ ID NO:22; 7) CABF-3 as defined in SEQ ID NO:23 and 8) SFL-1 as defined in SEQ ID NO:24.

19. The method of Claim 17, wherein the TFSRP coding nucleic acid is selected from the group consisting of 1) APS-2 as defined in SEQ ID NO:9; 2) ZF-2 as defined in SEQ ID NO:10; 3) ZF-3 as defined in SEQ ID NO:11; 4) ZF-4 as defined in SEQ ID NO:12; 5) ZF-5 as defined in SEQ ID NO:13; 6) MYB-1 as defined in SEQ ID NO:14; 7) CABF-3 as defined in SEQ ID NO:15 and 8) SFL-1 as defined in SEQ ID NO:16.

20. The method of Claim 17, wherein the TFSRP coding nucleic acid hybridizes under stringent conditions to a sequence of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16.

21. A method of modifying stress tolerance of a plant comprising, modifying the expression of a Transcription Factor Stress-Related Protein (TFSRP) in the plant, wherein the TFSRP is selected from the group consisting of 1) a AP2 Similar-2 protein (APS-2); 2) a Zinc-Finger Factor-2 protein (ZF-2); 3) a Zinc-Finger Factor-3 protein (ZF-3); 4) a Zinc-Finger Factor-4 protein (ZF-4); 5) a Zinc-Finger Factor-5 protein (ZF-5); 6) a MYB-1 protein (MYB-1); 7) a CAAT-Box Binding Factor protein-3 (CABF-3); 8) a Sigma Factor Like protein (SFL-1); and orthologs thereof.

22. The method of Claim 21, wherein the TFSRP is selected from the group consisting of 1) APS-2 as defined in SEQ ID NO:17; 2) ZF-2 as defined in SEQ ID NO:18; 3) ZF-3 as defined in SEQ ID NO:19; 4) ZF-4 as defined in SEQ ID NO:20; 5) ZF-5 as defined in SEQ ID NO:21; 6) MYB-1 as defined in SEQ ID NO:22; 7) CABF-3 as defined in SEQ ID NO:23 and 8) SFL-1 as defined in SEQ ID NO:24.

23. The method of Claim 21, wherein the TFSRP coding nucleic acid is selected from the group consisting of 1) APS-2 as defined in SEQ ID NO:9; 2) ZF-2 as defined in SEQ ID NO:10; 3) ZF-3 as defined in SEQ ID NO:11; 4) ZF-4 as defined in SEQ ID NO:12; 5) ZF-5 as defined in SEQ ID NO:13; 6) MYB-1 as defined in SEQ ID NO:14; 7) CABF-3 as defined in SEQ ID NO:15 and 8) SFL-1 as defined in SEQ ID NO:16.

24. The method of Claim 21, wherein the TFSRP coding nucleic acid hybridizes under stringent conditions to a sequence of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 or SEQ ID NO:16.

25. The method of Claim 21, wherein the stress tolerance is increased.

26. The method of Claim 21, wherein the stress tolerance is decreased.

27. The method of Claim 21, wherein the plant is not transgenic.

28. The method of Claim 21, wherein the plant is transgenic.

29. The method of Claim 28, wherein the plant is transformed with a promoter that directs expression of the TFSRP.

30. The method of Claim 29, wherein the promoter is tissue specific.

31. The method of Claim 29, wherein the promoter is developmentally regulated.

32. The method of Claim 21, wherein TFSRP expression is modified by administration of an antisense molecule that inhibits expression of TFSRP.

FIGURE 1A

Nucleotide sequence of the partial APS-2 from *Physcomitrella patens* (SEQ ID NO:1)

TCAAGCCACTCATCCGAGCATAGAACATCACAACCCACCTTGATGATCATTCTCTCA

GCCGACCAGCGTCAATTACGCTGCGTATCGCTCTAGCTTGAGGAAGGCACCCTCGCC

CTCTTCGCCGCGGAAGTAGCCCTCTGCTTCACGAGGGCGGCAAAACTCTCCCAAGGC

AGTTCCGGGGGGATGGGATATAGCTGCAGCTGCTGTGGGGAATCCTCAAAATTGTA

CGGGATCTTCTTCTTTGTGTAGAAGATGCCAACATCGTAGGCCCGGGCAGCTTCTTC

CGGAGTTTCATATGTTCCCAGCCATATCTTACGTTTCTGAGATGTGGGTCGAATTTCT

GTCACCCATTTGTTTAGCTCGGGCCGGTGCCGAACCCCCCTAAAACTGGTCGTATCG

CCAGTGTTGCTAGCAGAAACTTCTTCGGTATCCCATGCCGATGGGGCCTTATTTAAA

TCAATATTCCGAAATTTAAAGGCATTCCGACCGCTAGTGTCTTTCGCCGCTAACCGC

ATTCCCTTGGATTCTTCTTCCAAACTAGATTCAGACTTGCTCTCCTGCCAACTTCTTTT

TTCACTTTCGGGGATTCTATTTTAGTCGTTAACTGCAACGCCTGTTCTTTGACCTTGC

CACCACAAGGATCCCACTTCTTTGTTTTGGGCTTCCCCTGTTCAATAATGCTGGAAAT

TGTCAAATTCATGAACTACCCAATTGCAACCCCTCCCACCGGGATGGATTGATCGCC

AAAATTTCGTAGTAACTTAACTTTCATACAACAACTTGAGTTCCTTCGCTATTAGGG

ACACGTGGCAGAAACTTGGACGTGCAAGCGTATGTACTCATCAGAGTTTGACAGCG

CATAAAATCATATAAAAAGTCTTGAAGAAGCGTTGTTTAATTCATGGGTAACCACGA

GTTACGCGGAGCGTCGGCAGCAAGGAGAGGACGACCAGGCGGCAAGAAGATGCGT

CGGCAAGAGCTCGTGC

FIGURE 1B

Nucleotide sequence of the partial ZF-2 from *Physcomitrella patens* (SEQ ID NO:2)

TTTTTTTTGGCGAAAATGGGTAAAAATTTCCGTGGCCGTAAAATCAGTTGTGGCGCT

TGCCTTGCAATAAGCTGGTTATCGTAAAATGGCAATTACCTTGATATGTTCACTAGG

TTCGTGTCAGTGAGATGCCTTGCAGAAGGCGATTTCCCGTTTATTTACAACTCTACAT

GTTACTGAAGCACTGTGGCATTCAATCTCCTAACCTAGAGATGCTTAACTGCTTCGT

GCATGATCTATATTACATCTTGAGCCTCAGACTGTGGCGTCTTATTGCACATCTAGCG

CTATATATTACACTACGGTACACCATCGGAAGAAGTGAAGAAGGAATAGCTACTAT

TTTGCATCTCCAGTGTCAAATGTGATGCTGACCTGACCTAATCGGCAGAGACGCAGA

TCTCCAATGCTCACACTTCACATCCTAAAACGCACTTGCAGTGTACCCTGCTCTCAAT

TCCCATTGCAATTCCACATCTGGGATCTAAGGGCATGCTTTGTGGGCACTAAAGCGC

CGATTTCCTCCTCAATGCGAAGGTTGACCTTACTGAGGAAAGTGCGCATCGTAATCG

GCGATGTAGGCTTAGGGATTTCGCCGGGTAACACTCAACGCAATCGTGACGTTCGCT

AGATAGCTTGGTTGATATCAGGTGCAAATCCCACAAAATCCTTTGCATGGTCGGGCA

TTTATGTCTACCAGGGAAGTATCAGTTTTCTTCCAGAATAAAATTTCCACTTTAACAG

CACCTGCTCACGAAATCCTCAGGCATGTGGTGGTGGTGGACGGGGTGAAGAAGAAG

GCCCGCCCTCGTCAACACCATCTTCTCCAGTTTGGGGCGACACCACACTCTTCCCTC

GGCTCAACAAACGTCGGAAGCTCGCTGAAGCACGCGCCATCGGCGACAAAACATTC

GACGAATTGCTGACAGCCGCTGGACCAGACCTCGGAATGTCGATAGTAACCTGAAA

CGGCGCACGAATACTGGACGCCGCCCTCGCTTCTGCACTCCCTCCTGCACCTGCACT

GCTCATCTGCGCATGATTCCCCCAGAATAACACATTGGACGGGATTCCGGACGGTGT

CTCGTAATCTTTTATGTTTTGCTGCTGATCAACCGCAACGGCCGTTTCCGACGTACCA

CTTCGGCGATGATTCTCCCCGCCCTGATCCTCCGCAGTGCGGGGCAAACTATTGCCT

CTTGGGGAACTATTCAACGCCGGCAACTGTCCCCGGCTCCGTTGGCTCCTCCTGGAA

GCTCGCATGGCCGCCATGAACGGCGCTCCTACGTCACCCATAACGGGCGCTTCCATC

FIGURE 1B (Continued)

TGAGGAGGCTCGCTTATCTGCATCACGGTGGCGGCCTCGGTCTTCTTCTCAGTTTCAT

CAGCTCCCACACTCGTGCGACACAACGGGCATGTCGAGTGCGAGTGCAACCACATG

TCAATACAATCCAAGTGGAAACTATGGTCACACTTCGGCAACGTGCGGCCTTTCTCA

CCCAACTCAAATTCTTCCAAACAAACCGCGCACTCGAACCCACGCTTTGCCCTCTCA

CCGTCGAATTCGAAAGTGGGCAGAGCTTCAATAACAGCCCGCTCAAGCCCCACTGC

CTGCGTCACCGGAGTAGCGTTCACAGGGACAGTGTAACGGCGTCTTCGCCATGATAA

CGTACGCAAGGTGCCATCGCTTGCGACGATCTGGTGC

FIGURE 1C

Nucleotide sequence of the partial ZF-3 from *Physcomitrella patens* (SEQ ID NO:3)

GCACCAGCGCTTTTAAACAATCAAATATCTGAGCAGGTTGATGGCGAAGATTCAGAT

GTATCAAGAAGTGGCGCTAGTGATCAATCTGGGCATGAAAGACCCCTTGACGTTTTA

CGCAAGGTGAAAGGAAATGATGCTTGTGCCGACTGCGGTGCTGCTGATCCCGATTG

GGCTTCGCTGAATCTTGGGATTCTTCTGTGTATTGAGTGCTCAGGAGTACACAGAAA

CATGAGCGTTCAGATTTCTAAGGTCCGTTCGTTGACGTTAGATGTCAAAGTTTGGGA

GCCTTCTGTAATGAGCTATTTTCAATCTGTCGGAAACTCCTACGCTAATTCTATATGG

GAAGAGCTTTTGAATCCCAAGTCCTCAGAGGAGTCAAGTGAGAGAAACGTTAATGA

CGAGGGACAATCGGGCGTTTTAAGTGCTAGCAGAGCAAGGCCAAGACCTAGAGACC

CCATACCTATCAAAGAAAGATTTATCAATGCAAAGTATGTGGAGAAAA

FIGURE 1D

Nucleotide sequence of the partial ZF-4 from *Physcomitrella patens* (SEQ ID NO:4)

GCACGAGCTGCCCCATTCGAGCCCACTCGCACGAGAAGATACGAGCCGCGCTTTGG

CCGAGTGGTCGTAAGTAGAAGTAAAGGTCCGCGGCCGCTGCGGTCTGTGAAATTCTC

TCGCACGGAGAGAAGCGTGTTCTGCTGGTTTCTTCGCACAGAGACTTCTCCTGCACC

TTTTCTTCTTCCTCTACATCGTCTCCTGCGACGACTACATTGTGTGGGAGCAGTGGCA

ACCTTCCTGGCCACCCCGGGCTTCCTCTCAGTCAGTGGCTCACGTCTCCCAGCTAGG

CCTCCCATCGCGTCTTGCCGGCTCAATCGGGTGTCTTGCTCTGTTTTTTAACCTCTCTC

CCTTCCGGCCCTCTTATTCCTCTCCAGTCACTTCCGCCGGATCGCGACTTTTGTACCC

ATTTGGGGGTTGGGTGTTATAAGTTTGCCCTCAGGGTGTGAGTTGCTTTGTGTGTCTT

TTGTAGTAGTACTTTGCTTGTTGGGTGCGGAAGGGAACCTTTGAGAAGTCGACCCAT

TCTCTAGTTTTGCACCAGTCCCGCTTAGTGTGTGTGTCATTAGTGTTGGTTGCAAGTC

TGAAGCCTTGAGCGAGATTTGCAGGATTTTCTCATACGCTTCTGATTAGGAAAGATA

CATCCTTATTAGTCTGTTAAAGATGGCCACCGAGCGTGTGTCTCAGGAGACGACCTC

GCAGGCCCCTGAGGGTCCAGTTATGTGCAAGAACCTTTGCGGCTTCTTCGGCAGCCA

AGCTACCATGGGGTTGTGCTCGAAGTGCTACCGAGAGACAGTCATGCAGCGAAGAT

GACGGCTTTAGCTGAGCAAGCCACTCAGGCTGCTCAGGCGACATCTGCCACAGCTG

CTGCTGTTCAGCCCCCCGCTCCTGTACATGAGACCAAGCTCACATGCGAGGTTGAGA

GAACAATGATTGTGCCGCATCAATCTTCCAGCTATCAACAAGACCTGGTTACCCCCG

CTGCAGCTGCCCCTCAGGCAGTGAAGTCCTCTATCGCAGCTCCCTCTAGACCCGAGC

CCAATCGATGCGGATCTTGCAGGAAGCGTGTTGGATTGACAGGATTTAAGTGTCGCT

GTGGCAACCTCTACTGCGCTTTACATCGGTACTCGGACAAACACACTTGCACATATG

ACTACAAAGCCGCAGGGCAGGAAGCGATTGCGAAAGCTAATCCTCTTGTCGTGGCC

GAGAAGGTTGTCAAGTTTTGATGAGCATCCGTTAAGCTTTTCTGCCGACGATTTAGG

CTTCATACATTGAGTAACTCTACATCTTTCTTCTTTATCGAGAGAGCGAGTCGCATCA

FIGURE 1D (Continued)

AGATGAAGTCGAGGGGTGCGCGTCGGTTTTGGGGAGAGGGGATTTCTTTCCCCTTTC

CCCCCTTGGCGGCATCGTGTTTTATGTGTACAGAAGTAGGTTAGGACAAGATAGAAT

CATATGCCAGATCAATTGATAGTCCTCTTTAAGGAGGACACTTATTACACAATAAAA

AATCCTGGGTAATGCATGCCTTGATTGTGTTGTTTTTTCCTCGTGC

FIGURE 1E

Nucleotide sequence of the partial ZF-5 from *Physcomitrella patens* (SEQ ID NO:5)

GCGTGGGGCGTCTACACTAGTTTATCCCCGGGCTGAGGAATTCGGCACCAGATTTGT

CAATCAAAAGAAGTTAGTTGCGGGTGATGCTATTGTATTTCTTCGCATCGCATCTGG

CGAACTTTGTGTCGGCGTGCGCCGTTCAATGAGGGGTGTCAGCAACGGAGAATCCTC

ATCTTGGCACTCCTCAATCAGTAATGCTTCAACGATTCGGCCATCTCGATGGGAGGT

GAAGGGCACAGAAAGTTTCTCGGACTTTTTAGGTGGCGTTGGTGATAATGGGTACGC

ACTGAATAGCTCAATTCGGTCTGAAAACCAGGGCTCTCCAACAACGAGTAGCTTTGC

ACGGGACCGTGCTCGTGTTACTGCGAAGTCCGTTCTAGAAGCTGCTGCACTCGCCGT

CTCCGGTGAACGTTTTGAGGTTGTGTATTATCCTCGTGCTAGCACAGCTGAGTTCTGT

GTCAAAGCTGGGCTTGTTAAACGTGCGCTAGAGCAATCGTGGTACGCTGGAATGCG

CTTCAAAATGGCATTTGAAACTGAAGACTCCTCGAGGATAAGCTGGTTTATGGGAAC

TATTGCTGCTGTTCAAGCAGCAGATCCAGTAACTTGTGGCCTAGTTCTCCATGGCGG

GTCTGCAGGTCACCTTGGGATGAGCGGACCTATTGCAGGAGTGATCGTGTAGCCATG

GAGTA

FIGURE 1F

Nucleotide sequence of the partial MYB-1 from *Physcomitrella patens* (SEQ ID NO:6)

GCACCAGTGTTCCCTTTCATATGCTCAGCATGTCCGCCAATGAGCGCGCCTGTTGTGT

ACAGTCTGTGGAGAGCTGTAGAAAATTCAATTCCGATTTCAAAATATCCAGCGACGA

TGACACGGAACATGGGAGTTTGGAGGACGACATGAAGGAGTTGAACGAAGACATG

GAAATTCCCTTAGGTCGAGATGGCGAGGGTATGCAGTCAAAGCAGTGCCCGCGCGG

CCACTGGCGTCCAGCGGAAGACGACAAGCTGCGAGAACTAGTGTCCCAGTTTGGAC

CTCAAAACTGGAATCTCATAGCAGAGAAACTTCAGGGTCGATCAGGGAAAAGCTGC

AGGCTACGGTGGTTCAATCAGCTGGACCCTCGCATCAACCGGCACCCATTCTCGGAA

GAAGAGGAAGAGCGGCTGCTTATAGCACACAAGCGCTACGGCAACAAGTGGGCATT

GATCGCGCGCCTCTTTCCGGGCCGCACAGACAACGCGGTGAAGAATCACTGGCCC

FIGURE 1G

Nucleotide sequence of the partial CABF-3 from *Physcomitrella patens* (SEQ ID NO:7)

GCACCAGGTCTTCGACTTTGCTTCAGCACGCGCGCGTTGTGGTCGATCTCTCGCTGG

AGCAACAGGTTGTCTTGTCGCTGCCATTGCTAAAGCCATTCTTACTTCTAGCACTTCT

CGGAGGTTATTGATTTCTCGCAAATTGCTCTTCCACCTGCCCTCTTTCGTGAGGGAGT

TCGAAGCTGAAAAGTAATGAGCTGAAGATTAAGGTCTTTTACGAGTGAACAGCGAG

CACACAGCTAGCAACTCTTTCGGAGAATACTCCAGGCGAAATTGGTCGGATGGCCG

ATAGCTACGGTCACAACGCAGGTTCACCAGAGAGCAGCCCGCATTCTGATAACGAG

TCCGGGGGTCATTACCGAGACCAGGATGCTTCTGTACGGGAACAGGATCGGTTCCTG

CCCATCGCGAACGTGAGCCGAATCATGAAGAAGGCGTTGCCGTCTAATGCAAAAAT

TTCGAAGGACGCGAAAGAGACTGTGCAGGAGTGTGTGTCCGAGTTCATCAGCTTCAT

CACTGGTGAGGCGTCAGATAAGTGC

FIGURE 1H

Nucleotide sequence of the partial SFL-1 from *Physcomitrella patens* (SEQ ID NO:8)

GCACGAGTTTTCTTGTGTCAAAGCAGCAGAAGAAATCCACTTCTGGTAGTATTCAAA

CATAAAAGAATGGAAACTTATGTAACAGTCTACTTTCTGATCGAAACATTACCAAAT

GCCTTTTTCCTGGTTTGGTAGGTACTATCAATCAGCAGCAATTAAATAGCGTCAGAT

TTCACATCTAAGTACTCTCGTAGAATGCTGTTCCGGCTGGGTTGCCTCAGCTTGCGCA

TCGCTTTTGCCTCAATTTGTCTTATCCTTTCCCGAGTAACTTTAAAGATTTGACCTATT

TCTTCTAAAGTCTTGGACCGCCCATCGTCCAATCCAAAACGCAGTCTTAGCACCTCC

CTCTCTCTTGGGTTCAATGTGCGTAGAACGCCCTCTATATCTTGTCGCATCAATTGCT

TTACGATTGCGTCCTCAGGTGAATCCACATCTGTGTCTGCGACAAGTTCCCCAAGTG

TANTGTCCCCATCTTTGCCAATGGGCCGCTCCATCGAACCTGGTGCCTTTGCTGATTT

CACTACAGATTTCAGTTTCTCAACAGTCAAGCCCACTAGCTCAGCCACTTCCTCGTT

ACGTGCTTCCCGCCCATGCTCCTG

FIGURE 2A

Nucleotide sequence of the full-length APS-2 from *Physcomitrella patens* (SEQ ID NO:9)

GCGATATCGGAAGAAGAACCAAGGGAATGCGGTTAGCGGCGAAAGACACTAGCGG

TCGGAATGCCTTTAAATTTCGGAATATTGATTTAAATAAGGCCCCATCGGCATGGGA

TACCGAAGAAGTTTCTGCTAGCAACACTGGCGATACGACCAGTTTTAGGGGGGTTCG

GCACCGGCCCGAGCTAAACAAATGGGTGACAGAAATTCGACCCACATCTCAGAAAC

GTAAGATATGGCTGGGAACATATGAAACTCCGGAAGAAGCTGCCCGGGCCTACGAT

GTTGGCATCTTCTACACAAAGAAGAAGATCCCGTACAATTTTGAGGATTCCCCACAG

CAGCTGCAGCTATATCCCATCCCCCCGGAACTGCCTTGGGAGAGTTTTGCCGCCCTC

GTGAAGCAGAGGGCTACTTCCGCGGCGAAGAGGGCGAGGGTGCCTTCCTCAAGCTA

GAGCGATACGCAGCGTAATTGACGCTGGTCGGCTGAGAGAATGATCATCAAGGTGG

GTTGTGATGTTCTATGCTCGGATGAGTGGCTTGAAGGGTTCTGGTTCCAACCATGAG

AGCATGACGCGAGTCCCACACGGATGGAGCTTGTGAATGGAGTGGTAGACTGTAGA

TGGTTTTTGTAACGGCTTGAGTAATAACGGAAGCTTCATGGCTTGAATGACCAGCCA

TGGTGGTGTGCAAGTGAAGATCGCTGCTTGTGTGAAGGTTTCCATCTTTCCCATCCCC

GTCTTCCACTTTGCTACACGTTGCTAGTGTCACTTGAACAATTCATTCATGGACCCTG

CTCTCCTTTCCCCTGTTACGAAGTTCTTATGGTAGAGTTCACCGAACGCAAGCTGTCT

AGGAAGTTGACAGTTTGTGGGAGCCAAAAACTCTACTTGAGCTACTGTGTGCACGCC

TTCTGAGTCCTCCAGCGAGGAGCCTGTATATTATTGGATGGTGCAGGATGGGTCGCT

TGGTGCCTTTCTCTTTTTCCTTTTCCTCTTTTTGTAAATGGTTTTCCTTCTATGAATATG

TGAAGCTCCTCCCACGGAAGCATAGAGCTCGC

FIGURE 2B

Nucleotide sequence of the full-length ZF-2 variant from *Physcomitrella patens* (SEQ ID NO:10)

ATCCCGGGATCAGGAAGCTGTCAAGGAAGAGATGGAAATCTTGCTCCATACAATTA

CTACGGGCCGCCACCGGGCAGTAACAATTATGTCGTCAACAGCAAGATTATGGTCGT

GGCTGTCGCGGTTCTCTTCGCTGTCGTCCTCTTCATCCTCTGCCTCCACATCTACGCC

AAGTGGTTCTGGCGCAATCAAGGTGCCATCGTCGCAAGCGATGGCACCTTGCGTACG

TTATCATGGCGAAGACGCCGTTACACTGTCCCTGTGAACGCTACTCCGGTGACGCAG

GCAGTGGGGCTTGAGCGGGCTGTTATTGAAGCTCTGCCCACTTTCGAATTCGACGGT

GAGAGGGCAAAGCGTGTGTTCGAGTGCGCGGTTTGTTTGGAAGAATTTGAGTTGGGT

GAGAAAGGCCGCACGTTGCCGAAGTGTGACCATAGTTTCCACTTGGATTGTATTGAC

ATGTGGTTGCACTCGCACTCGACATGCCCGTTGTGTCGCACGAGTGTGGGAGCTGAT

GAAACTGAGAAGAAGACCGAGGCCGCCACCGTGATGCAGATAAGCGAGCCTCCTCA

GATGGAAGCGCCCGTTATGGGTGACGTAGGAGCGCCGTTCATGGCGGCCATGCGAG

CTTCCAGGAGGAGCCAACGGAGCCGGGGACAGTTGCCGGCGTTGAATAGTTCCCCA

AGAGGCAATAGTTTGCCCCGCACTGCGGAGGATCAGGGCGGGGAGAATCATCGCCG

AAGTGGTACGTCGGAAACGGCCGTTGCGGTTGATCAGCAGCAAAACATAAAAGATT

ACGAGACACCGTCCGGAATCCCGTCCAATGTGTTATTCTGGGGGAATCATGCGCAGA

TGAGCAGTGCAGGTGCAGGAGGGAGTGCAGAAGCGAGGGCGGCGTCCAGTATTCGT

GCGCCGTTTCAGGTTACTATCGACATTCCGAGGTCTGGTCCAGCGGCTGTCAGCAAT

TCGTCGAATGTTTTGTCGCCGATGGCGCGTGCTTCAGCGAGCTTCCGACGTTTGTTGA

GCCGAGGGAAGAGTGTGGTGTCGCCCCAAACTGGAGAAGATGGTGTTGACGAGGGC

GGGCCTTCTTCTTCACCCCGTCCACCACCACCACATGCCTGAGGATTTCGTGAGCAG

GTGCTGTTAAAGTGGAAATTTTATTCTGGAAGAAAACTGATACTTCCCTGGTAGACA

TAAATGCCCGACCATGCAAAGGATTTTGTGGGATTTGCACCTGATATCAACCAAGCT

ATCTAGCGAACGTCACGATTGCGTTGAGTGTTACCCGGCGAAATCCCTAAGCCTACA

FIGURE 2B (Continued)

TCGCCGATTACGATGCGCACTTTCCTCAGTAAGGTCAACCTTCGCATTGAGGAGGAA

ATCGGCGCTTTAGTGCCCACAAAGCATGCCCTTAGATCCCAGATGTGGAATTGCAAT

GGGAATTGAGAGCAGGGTACACTGCAAGTGCGTTTTAGGATGTGAAGTGTGAGCAT

TGGAGATCTGCGTCTCTGCCGATTAGGTCAGGTCAGCATCACATTTGACACTGGAGA

TGCAAAATAGTAGCTATTCCTTCTTCACTTCTTCCGATGGTGTACCGTAGTGTAATAT

ATAGCGCTAGATGTGCAATAAGACGCCACAGTCTGAGGCTCAAGATGTAATATAGA

TCATGCACGAAGCAGTTAAGCATCTCTAGGTTAGGAGATTGAATGCCACAGTGCTTC

AGTAACATGTAGAGTTGTAAATAAACGGGAAATCGCCTTCTGCAAGGCATCTCACTG

ACACGAACCTAGTGAACATATCAAGGTAATTGCCATTTTACGATAACCAGCTTATTG

CAAGGCAAGCGCCAGAGCTCGC

FIGURE 2C

Nucleotide sequence of the full-length ZF-3 from *Physcomitrella patens* (SEQ ID NO:11)

ATCCCGGGAGGAGGACTTGCGGAATGCAAAATCACAATTTGAGCAGGCTCGATTCA

ATTTGATGACAGCACTTACCAATAGTGAGGCAAAAAAGAAGTTCGAGTTCCTTGAA

GCCGTGAGTGGTACAATGGATGCACATCTCAGGTACTTCAAGCAGGGCTATGAGTTG

CTACATCAAATGGAACCTTACATCCATCAGGTGTTAACATATGCTCAACAGTCCAGA

GAAAGGGCCAACTACGAGCAAGCAGCACTTGCAGATCGTATGCAGGAGTACAGGCA

GGAAGTTGAGAGAGAGAGCCAAAGGTCGATTGATTTTGACAGCTCTTCTGGAGATG

GTATTCAAGGTGTTGGCCGCAGTTCACATAAGATGATTGAGGCAGTCATGCAATCGA

CCCCAAAAGGGCAGATCCAGACTCTTAAGCAGGGATACCTGTTAAAGCGTTCAACA

AATTTGCGAGGTGACTGGAAGCGGAGGTTTTTTGTGTTGGATAGCAGAGGAATGCTG

TATTATTATCGGAAACAGTGGGGCAAGCCTACAGACGAGAAAAATGTAGCACATCA

CACTGTGAATCTGCTGACGTCTACAATCAAGATAGACGCAGAACAATCAGATCTTCG

TTTCTGCTTTCGGATTATTTCTCCAGCTAAAAGTTATACCCTCCAGGCAGAAAATGCC

ATTGACAGAATGGATTGGATGGACAAAATTACAGGGGTGATTTCGTCGCTTTTAAAC

AATCAAATATCTGAGCAGGTTGATGGCGAAGATTCAGATGTATCAAGAAGTGGCGC

TAGTGATCAATCTGGGCATGAAAGACCCCTTGACGTTTTACGCAAGGTGAAAGGAA

ATGATGCTTGTGCCGACTGCGGTGCTGCTGATCCCGATTGGGCTTCGCTGAATCTTG

GGATTCTTCTGTGTATTGAGTGCTCAGGAGTACACAGAAACATGAGCGTTCAGATTT

CTAAGGTCCGTTCGTTGACGTTAGATGTCAAAGTTTGGGAGCCTTCTGTAATGAGCT

ATTTTCAATCTGTCGGAAACTCCTACGCTAATTCTATATGGGAAGAGCTTTTGAATCC

CAAGTCCTCAGAGGAGTCAAGTGAGAGAAACGTTAATGACGAGGGACAATCGGGCG

TTTTAAGTGCTAGCAGAGCAAGGCCAAGACCTAGAGACCCCATACCTATCAAAGAA

AGATTTATCAATGCAAAGTATGTGGAGAAAAAATTTGTCCAAAAGTTGAAGGTGGA

TTCTCGAGGCCCGTCAGTGACACGGCAGATCTGGGATGCTGTCCAGAACAAAAAAG

FIGURE 2C (Continued)

TGCAGCTTGCTCTTCGTCTTCTTATCACTGCTGATGCTAACGCCAACACAACCTTCGA

GCAAGTAATGGGTGGTACCGAGTCTTCGTGGTCGTCTCCACTTGCAAGCCTCGCTGG

AGCTCTCTTACGAAAGAACTCTCTCAGTGCCTCTCAGAGTGGTCGCAGGAACTGGAG

CGTACCTTCACTATTGTCTTCTCCAGACGATCCGGGGTCCCGTTCAGGAGCTTTAAGC

CCTGTTTCGAGAAGTCCTGATGCAGCAGGCAGCGGAGGGATTGATGAGAAAGATTT

GCGGGGCTGCAGTTTGCTCCATGTTGCCTGCCAAATCGGAGATATTAGCCTGATCGA

GCTGCTACTTCAATACGGGGCGCAAATCAATTGTGTGGATACCCTGGGTCGAACTCC

TCTTCATCACTGTGTCTTGTGCGGCAACAATTCGTGTGCAAAGCTCCTGCTCACAAG

AGGGGCGAAGGCGGGTGCCGTAGACAAAGAGGGAAAAACTCCGCTGGAGTGTGCA

GTGGAGAAGCTAGGTGCTATCACGGATGAAGAATTGTTCATAATGCTTTCTGAAACC

AGTAGATGACACCACATTTGTGCCTGAGTTGCTTTGTGTATAAATCTCAACATCAAC

TTGTTTCCTAGCACCTGTAAGGCTAGTTTGTTTGGGTAGTTTGCATTCTTGTTCTACC

GTTTTATCTTCCCATTACGTCAGCATAAGTAGAGAGTGGAAGCAGGTGGATATCGC

FIGURE 2D

Nucleotide sequence of the full-length ZF-4 from *Physcomitrella patens* (SEQ ID NO:12)

ATCCCGGGCACCAGTCCCGCTTAGTGTGTGTGTCATTAGTGTTGGTTGCAAGTCTGA

AGCCTTGAGCGAGATTTGCAGGATTTTCTCATACGCTTCTGATTAGGAAAGATACAC

CCTTATTAGTCTGTTAAAGATGGCCACCGAGCGTGTGTCTCAGGAGACGACCTCGCA

GGCCCCTGAGGGTCCAGTTATGTGCAAGAACCTTTGCGGCTTCTTCGGCAGCCAAGC

TACCATGGGGTTGTGCTCGAAGTGCTACCGAGAGACAGTCATGCAAGCGAAGATGA

CGGCTTTAGCTGAGCAAGCCACTCAGGCTGCTCAGGCGACATCTGCCACAGCTGCTG

CTGTTCAGCCCCCCGCTCCTGTACATGAGACCAAGCTCACATGCGAGGTTGAGAGAA

CAATGATTGTGCCGCATCAATCTTCCAGCTATCAACAAGACCTGGTTACCCCCGCTG

CAGCTGCCCCTCAGGCAGTGAAGTCCTCTATCGCAGCTCCCTCTAGACCCGAGCCCA

ATCGATGCGGATCTTGCAGGAAGCGTGTTGGATTGACAGGATTTAAGTGTCGCTGTG

GCAACCTCTACTGCGCTTTACATCGGTACTCGGACAAACACACTTGCACATATGACT

ACAAAGCCGCAGGGCAGGAAGCGATTGCGAAAGCTAATCCTCTTGTCGTGGCCGAG

AAGGTTGTCAAGTTTTGATGAGCATCCGTTAAGCTTTTCTGCCGACGATTTAGGCTTC

ATACATTGAGTAACTCTACATCTTTCTTCTTTATCGAGAGAGCGAGTCGCATCAAGA

GCTCGCC

FIGURE 2E

Nucleotide sequence of the full-length ZF-5 from *Physcomitrella patens* (SEQ ID NO:13)

ATCCCGGGTATCGATCTGGAGCCCGTTGCAAACTCAATGGTGTATTTTATAGGGCAA

AAGTCTGATCTATATGGAATGCATCCTCTCAGAGTTGCAAATCATGGACTGCATGTC

ACTCTGGGTTATTCTCGATCACCTAGCTTTGCTGGAGTTCAAATTGGTGAGTACGAG

TATTATGAGTGATCTCGAGTTTATGGTCCCCTTCTTTCATGATCAAGGGTAATTTATA

TCAAGGGTGTATATGAGAGATACGCACTTATTGAGTGGACCTTTTCTCATACTGCAT

TTACACCCCTGTCAGTTGCAGCATCCTGGTTTGGAATGCCGGGTCCAGTCCCTCTATT

ATCCATGAGTGTAAAATCGGAGAGTCTCGATGACATTGGAGGTCACGAGAAAAAAT

CTGTAACTGGGTCGGAAGTGGGTGGCCTCGATGCTCAGCTGTGGCATGCCTGTGCTG

GGGGTATGGTTCAACTGCCTCATGTGGGTGCTAAGGTTGTCTATTTTCCCCAAGGCC

ATGGCGAACAAGCTGCTTCAACTCCCGAGTTCCCCCGCACTTTGGTTCCAAATGGAA

GTGTTCCCTGCCGAGTTGTGTCAGTTAACTTTCTGGCTGATACAGAAACAGACGAGG

TATTTGCTCGTATTTGCCTGCAGCCTGAGATTGGCTCCTCCGCTCAGGATTTAACAGA

TGATTCTCTTGCGTCTCCGCCTCTAGAGAAACCAGCTTCATTTGCCAAAACGCTCACT

CAAAGTGATGCAAACAACGGTGGAGGCTTTTCAATACCTCGTTATTGTGCTGAAACT

ATTTTCCCACCTCTCGATTACTGTATCGATCCTCCTGTTCAAACTGTTCTTGCAAAAG

ATGTCCATGGAGAGGTGTGGAAATTTCGTCACATTTACAGGGGGACTCCACGTCGAC

ATTTGTTAACCACAGGATGGAGCACATTTGTCAATCAAAAGAAGTTAGTTGCGGGTG

ATGCTATTGTATTTCTTCGCATCGCATCTGGCGAACTTTGTGTCGGCGTGCGCCGTTC

AATGAGGGGTGTCAGCAACGGAGAATCCTCATCTTGGCACTCCTCAATCAGTAATGC

TTCAACGATTCGGCCATCTCGATGGGAGGTGAAGGGCACAGAAAGTTTCTCGGACTT

TTTAGGTGGCGTTGGTGATAATGGGTACGCACTGAATAGCTCAATTCGGTCTGAAAA

CCAGGGCTCTCCAACAACGAGTAGCTTTGCACGGGACCGTGCTCGTGTTACTGCGAA

GTCCGTTCTAGAAGCTGCTGCACTCGCCGTCTCCGGTGAACGTTTTGAGGTTGTGTAT

FIGURE 2E (Continued)

TATCCTCGTGCTAGCACAGCTGAGTTCTGTGTCAAAGCTGGGCTTGTTAAACGTGCG

CTAGAGCAATCGTGGTACGCTGGAATGCGCTTCAAAATGGCATTTGAAACTGAAGA

CTCCTCGAGGATAAGCTGGTTTATGGGAACTATTGCTGCTGTTCAAGCAGCAGATCC

AGTACTTTGGCCTAGTTCTCCATGGCGGGTTCTGCAGGTCACTTGGGATGAGCCGGA

CCTATTGCAGGGAGTGAATCGTGTAAGCCCATGGCAGTTAGAGCTTGTGGCGACACT

TCCTATGCAGCTGCCCCCTGTCTCTCTTCCCAAAAAGAAACTGCGCACTGTCCAGCC

TCAAGAGCTTCCACTTCAGCCCCCTGGACTGCTAAGCCTGCCGTTGGCAGGGACTAG

CAACTTTGGTGGGCACTTGGCCACCCCCTGGGGCAGCTCTGTTCTTTTGGATGACGC

CTCTGTTGGCATGCAGGGGGCCAGGCATGATCAATTCAACGGGCTTCCAACTGTGGA

TTTCCGAAATAGTAACTACAAACATCCTCGGGAGTTTTCTAGGGACAATCAGTACCA

GATTCAAGATCATCAAGTCTTCCATCCTAGACCTGTATTAAATGAGCCCCCTGCGAC

AAACACTGGCAACTACTTCTCTCTTTTACCTAGTCTCCAGCGACGGCCAGATATCTCT

CCTAGTATTCAGCCCTTAGCCTTCATGTCTGCTTCTGGAAGCTCACAGCTGGAGACTT

CTTCAACTAAGACAGCGGCCACCTCTTTTTTCCTATTTGGCCAATTCATTGACCCTTC

TTGCACCTCCAAACCTCAGCAGCGTTCCACAGTTATTAATAACGCTTCCGTTGCTGG

GGATGGTAAGCATCCTGGCACTAATAACTCATCCTCGGATAACAAATCAGAGGACA

AGGACAATTGTAGGGATGTTCAACCCATTCTGAATGGGATTGCTGTAAGATCTGGAT

TTCGAGCAGATATAGCCGCGAAGAAGTTTCAACAGAGCGACTCTGCACATCCCACG

GAAGCATCACGTGGAAGCCAAGTTAGCAGCTTACCGTGGTGGCAAACACAGGACGC

TCACAAGGATCAGGAATTCCATGGAGACAGTCAGACGCCTCATACTCCTGCATCTGG

TAGCCAATGAGGCTAAAGCTTGATCATAGCTCATAACCCTCTCACAGGACGTAATGG

GGGTGACAACATGCTAACAGAATTGCACGGTAAAGGAAAACTGTACTAGGCATGTT

ATATGGGAATTCGGATCGCTTCTTGCAATTAAACACGCTAGCGCCGTTTGGTGCCAA

FIGURE 2E (Continued)

TGTTATTCTGGCATTTGTTTTGTTTCCTTTGGAAACAAATTGCTATATTTCAAAGTCCT

TTGGAGGAGCTCGC

FIGURE 2F

Nucleotide sequence of the full-length MYB-1 from *Physcomitrella patens* (SEQ ID NO:14)

ATCCCGGGCTGTTGTGTACAGTCTGTGGAGAGCTGTAGAAAATTCAATTCCGATTTC

AAAATATCCAGCGACGATGACACGGAACATGGGAGTTTGGAGGACGACATGAAGG

AGTTGAACGAAGACATGGAAATTCCCTTAGGTCGAGATGGCGAGGGTATGCAGTCA

AAGCAGTGCCCGCGCGGCCACTGGCGTCCAGCGGAAGACGACAAGTTGCGAGAACT

AGTGTCCCAGTTTGGACCTCAAAACTGGAATCTCATAGCAGAGAAACTTCAGGGTCG

ATCAGGGAAAAGCTGCAGGCTACGGTGGTTCAATCAGCTGGACCCTCGCATCAACC

GGCACCCATTCTCGGAAGAAGAGGAAGAGCGGCTGCTTATAGCACACAAGCGCTAC

GGCAACAAGTGGGCATTGATCGCGCGCCTCTTTCCGGGCCGCACAGACAACGCGGT

GAAGAATCACTGGCACGTTGTGACGGCAAGACAGTCCCGTGAACGGACACGAACTT

ACGGCCGTATCAAAGGTCCGGTACATCGAAGAGGCAAGGGTAACCGTATCAATACC

TCCGCACTTGGAAATTACCATCACGATTCGAAGGGAGCTCTCACAGCCTGGATTGAG

TCGAAGTATGCGACAGTCGAGCAGTCTGCGGAAGGGCTCGCTAGGTCTCCTTGTACC

GGCAGAGGCTCTCCTCCTCTACCCACCGGTTTCAGTATACCGCAGATTTCCGGCGGC

GCCTTCCATCGACCGACAAACATGAGTACTAGTCCTCTTAGCGATGTGACTATCGAG

TCGCCAAAGTTTAGCAACTCCGAAAATGCGCAAATAATAACCGCGCCCGTCCTGCA

AAAGCCAATGGGAGATCCCAGGTCAGTATGCTTGCCGAATTCGACTGTTTCCGACAA

GCAGCAAGTGCTGCAGAGTAATTCCATCGACGGTCAGATCTCCTCCGGGCTCCAGAC

AAGCGCAATAGTAGCGCATGATGAGAAATCGGGCGTCATTTCAATGAATCATCAAG

CACCGGATATGTCCTGTGTTGGATTGAAGTCAAATTTTCAGGGGAGTCTCCATCCTG

GCGCTGTTAGATCTTCTTGGAATCAATCCCTTCCCCACTGTTTTGGCCACAGTAACAA

GTTGGTGGAGGAGTGCAGGAGTTCTACAGGCGCATGCACTGAACGCTCTGAGATTCT

GCAAGAACAGCATTCTAGCCTTCAGTTTAAATGCAGCACTGCGTACAATACTGGAAG

ATATCAACATGAAAACCTTTGTGGGCCAGCATTCTCGCAACAAGACACAGCGAACG

FIGURE 2F (Continued)

AGGTTGCGAATTTTTCTACGTTGGCATTCTCCGGCCTAGTGAAGCATCGCCAAGAGA

GGTTGTGCAAAGATAGTGGATCTGCTCTCAAGCTGGGACTATCATGGGTTACATCCG

ATAGCACTCTTGACTTGAGTGTTGCCAAAATGTCAGCATCGCAGCCAGAGCAGTCTG

CGCCGGTTGCATTCATTGATTTTCTAGGCGTGGGAGCGGCCTGAAGGCTGCGGAAAG

ATTTTAGCAAAGCTTTTATAACGTTTTTTTTGCACAGGGCTGTTTTTAGCTTGTATAC

CAGTAGGCACTTCTACTTCTTTTTCTTCTTTTCTTTTTCCCCTTTTCTTCTCCCCCCACT

TTCACCATTTCCGCCATAGCAGCCTTTGAATCACGTAATGGAACCTTTGGCGGCCTG

TATGAGGCACTTTTGGAGGCATCCCTGGACGAAGAATGGATCAAACCGTACTGCGG

ATGTCATGCTTTGAAGCTGCAATCCGAATTCAGTAGCATGCTGTGGATGACTCAAAA

GGAGTAGCTGCTTTGTGAAACTAATACTATACAGCGGATTTTGAAGACCCAAGTTTC

ATGTGGACAAGTCTGAAAAACTTATACGCCACCTCCATGGGCTTCTACGATGAATAT

GCGCTTTCGGCTTACACTGCGGCTCTTTTTTGCATATATATATACTTCCATTCAATTTT

ATTTGGAAATGTTTTGAATCTACCTTCTCGTACAAAACTGGGATCAGAAATCTTCCA

GGTTGTGGGTCGCAAGTTAACTCTGCAGATTGTGGCTGACACTTGGGCAATCGGCAA

CTTTATCTTTTTGTTTTTTACGCTTGAACGGACCTCAGCTGTACAGACACTCATCATG

TACATTCGATGCCATCTCTTGGCTTTCATGGAAGTTCAGATATCGGAAACTGTGACA

GAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTCTTGATGCACTGTGCGC

CGAGTTTGAGACTAGTTTAGAAAGATTGATGAAGCTAGCAGTAAATTGTTGGCCTCA

TCTGAAAGGTACGGCCTTTACTCCGTGAGCCCGGGAT

FIGURE 2G

Nucleotide sequence of the full-length CABF-3 from *Physcomitrella patens* (SEQ ID NO:15)

ATCCCGGGCAGCGAGCACACAGCTAGCAACTCTTTCGGAGAATACTCCAGGCGAAA

TTGGTCGGATGGCCGATAGCTACGGTCACAACGCAGGTTCACCAGAGAGCAGCCCG

CATTCTGATAACGAGTCCGGGGGTCATTACCGAGACCAGGATGCTTCTGTACGGGAA

CAGGATCGGTTCCTGCCCATCGCGAACGTGAGCCGAATCATGAAGAAGGCGTTGCC

GTCTAATGCAAAAATTTCGAAGGACGCGAAAGAGACTGTGCAGGAGTGTGTGTCCG

AGTTCATCAGCTTCATCACTGGTGAGGCGTCAGATAAGTGCCAGAGGGAGAAGAGA

AAGACGATCAACGGTGACGACTTGCTGTGGGCCATGAGTACACTTGGTTTCGAAGAT

TACGTGGAGCCTCTGAAGGTTTACCTACACAAATACCGGGAGCTAGAGGGAGAGAA

GGCTTCCACGGCCAAGGGTGGTGACCAGCAAGGAGGGAAAGAAGGGAGTCAAGGT

GTTATGGGGTCCATGGGTATGTCGGGCGGAATGAACGGTATGAACGGTACGATGAA

CGGGAATATGCATGGACATGGAATCCCGGTGTCGATGCAGATGCTGCAGCAGTCGT

ACGGACAGCAGGCACCTCCAGGGATGATGTATTCCCCTCATCAGATGATGCCGCAA

TACCAGATGCCAATGCAGTCTGGTGGAAACCAGCCTCGTGGAGTGTAGGAGGTTCC

ACGGCGAGGAGAATTTGAAATTGGGGAGATTGTCAACCGCGTGAGGGAGTGAGCTC

GC

FIGURE 2H

Nucleotide sequence of the full-length SFL-1- from *Physcomitrella patens* (SEQ ID NO:16)

ATCCCGGGCTCGGAAGGACTGTGCATTGTCGAGCGCTGAAGGTGGATGATGCTTTGG

TGACCGAGAGCGGTCTTATCAGTGAAGAAGGAGTTTCTCGTGCTGCAGCTGAGGAG

GCGATGACGTTAGCTTTAGCAGCTGCAAAGGCCGCCATGGAGGCTGCCTCGTACGCT

GATGCGATGCCGTGGAACAGGAGGAGTTTCCGACGGAATTTGATCTGCTGAGACTA

GAGAGGGCCAGGTTGAGCGATGTTGAGCATTCTTTTCGGGTTGAATTGGATACAGAG

GCTGCCATGATGGAGGCCGAGCAGAGTTATGTGCAGAAGCTAGAATCGTTGTTGGG

AGGTGTTTCCACGCTCGTCCGTGAGGAAGAGGAAACTGCATCCGTTTCAGAAGATG

AAGATGATTCAAACAGCTTACCTCAAATTCAAGTAGCCGTTAAATCGAAGCGGAAG

GGAGAGAGGAGGAAGAGGCGGGAGCGAGCGTTGGAAAGGGCAGAGAAGGTTGCCA

CCGATCTTGCATCAGCACCCCCTCTCCCAAAACCTAAGAAACCACAGCTTGCGGCGG

ATCCTTCAGACCCAGTCCGTGCATATTTGCGAGACATAGGAAGGACGAAGTTGCTAA

CAGCAAGAGAAGAAGTCGATCTCTCTCATCAAATTCAGGATCTTTTGAAGTTGGAGA

ATATCAAGTCTAACCTTGAGCGAGAGATAGGAAGGAATGCCACAATTGGAGAGTGG

AGTAGAGCGGTAGGAATGGAACAGAATGCATTTGAAGCGCGGCTTAAGAAGGGTCG

ATTCGCCAAGGACAAAATGGTGAATTCGAATTTGCGGTTGGTTGTCTCGATTGCGAA

AAACTACCAGGGCCGAGGCATGACTCTTCAAGATTTAATTCAGGAAGGGAGCATGG

GATTGGTGAGAGGAGCGGAGAAGTTCGACCCGACCAAGGGGTTTAAGTTCAGCACT

TACGCACATTGGTGGATTAGGCAGGCTGTAACGCGATCAATTGCGGATCAATCTAGG

ACTTTTCGTTTACCTATTCATTTATACGAAGTTATCTCACGTATCAACAAAGCAAAGC

GAATGCTGGTTCAGGAGCATGGGCGGGAAGCACGTAACGAGGAAGTGGCTGAGCTA

GTGGGCTTGACTGTTGAGAAACTGAAATCTGTAGTGAAATCAGCAAAGGCACCAGG

TTCGATGGAGCGGCCCATTGGCAAAGATGGGGACACTACACTTGGGGAACTTGTCG

CAGACACAGATGTGGATTCACCTGAGGACGCAATCGTAAAGCAATTGATGCGACAA

FIGURE 2H (Continued)

GATATAGAGGGCGTTCTACGCACATTGAACCCAAGAGAGAGGGAGGTGCTAAGACT

GCGTTTTGGATTGGACGATGGGCGGTCCAAGACTTTAGAAGAAATAGGTCAAATCTT

TAAAGCTACTCGGGAAAGGATAAGACAAATTGAGGCAAAAGCGATGCGCAAGCTG

AGGCAACCCAGCCGGAACAGCATTCTACGAGAGTACTTAGATGTGAAATCTGACGC

TATTTAATTGCTGCTGATTGATAGTACCTACCAAACCAGGAAAAAGGCATTTGGTAA

TGTTTCGATCAGAAAGTAGACTGTTACATAAGTTTCCATTCTTTTATGTTTGAATACT

ACCAGAAGTGGATTTCTTCTGCTGCGAGCTCGC

FIGURE 3A

Deduced amino acid sequence of APS-2 from *Physcomitrella patens* (SEQ ID NO:17)

MRLAAKDTSGRNAFKFRNIDLNKAPSAWDTEEVSASNTGDTTSFRGVRHRPELNKWVT

EIRPTSQKRKIWLGTYETPEEAARAYDVGIFYTKKKIPYNFEDSPQQLQLYPIPPELPWES

FAALVKQRATSAAKRARVPSSS*

FIGURE 3B

Deduced amino acid sequence of ZF-2 from *Physcomitrella patens* (SEQ ID NO:18)

MVVAVAVLFAVVLFILCLHIYAKWFWRNQGAIVASDGTLRTLSWRRRRYTVPVNATPV

TQAVGLERAVIEALPTFEFDGERAKRVFECAVCLEEFELGEKGRTLPKCDHSFHLDCIDM

WLHSHSTCPLCRTSVGADETEKKTEAATVMQISEPPQMEAPVMGDVGAPFMAAMRAS

RRSQRSRGQLPALNSSPRGNSLPRTAEDQGGENHRRSGTSETAVAVDQQQNIKDYETPS

GIPSNVLFWGNHAQMSSAGAGGSAEARAASSIRAPFQVTIDIPRSGPAAVSNSSNVLSPM

ARASASFRRLLSRGKSVVSPQTGEDGVDEGGPSSSPRPPPPHA*

FIGURE 3C

Deduced amino acid sequence of ZF-3 from *Physcomitrella patens* (SEQ ID NO:19)

MTALTNSEAKKKFEFLEAVSGTMDAHLRYFKQGYELLHQMEPYIHQVLTYAQQSRERA
NYEQAALADRMQEYRQEVERESQRSIDFDSSSGDGIQGVGRSSHKMIEAVMQSTPKGQI
QTLKQGYLLKRSTNLRGDWKRRFFVLDSRGMLYYYRKQWGKPTDEKNVAHHTVNLL
TSTIKIDAEQSDLRFCFRIISPAKSYTLQAENAIDRMDWMDKITGVISSLLNNQISEQVDGE
DSDVSRSGASDQSGHERPLDVLRKVKGNDACADCGAADPDWASLNLGILLCIECSGVH
RNMSVQISKVRSLTLDVKVWEPSVMSYFQSVGNSYANSIWEELLNPKSSEESSERNVND
EGQSGVLSASRARPRPRDPIPIKERFINAKYVEKKFVQKLKVDSRGPSVTRQIWDAVQNK
KVQLALRLLITADANANTTFEQVMGGTESSWSSPLASLAGALLRKNSLSASQSGRRNWS
VPSLLSSPDDPGSRSGALSPVSRSPDAAGSGGIDEKDLRGCSLLHVACQIGDISLIELLLQY
GAQINCVDTLGRTPLHHCVLCGNNSCAKLLLTRGAKAGAVDKEGKTPLECAVEKLGAI
TDEELFIMLSETSR*

FIGURE 3D

Deduced amino acid sequence of ZF-4 from *Physcomitrella patens* (SEQ ID NO:20)

MATERVSQETTSQAPEGPVMCKNLCGFFGSQATMGLCSKCYRETVMQAKMTALAEQA

TQAAQATSATAAAVQPPAPVHETKLTCEVERTMIVPHQSSSYQQDLVTPAAAAPQAVK

SSIAAPSRPEPNRCGSCRKRVGLTGFKCRCGNLYCALHRYSDKHTCTYDYKAAGQEAIA

KANPLVVAEKVVKF*

FIGURE 3E

Deduced amino acid sequence of ZF-5 from *Physcomitrella patens* (SEQ ID NO:21)

MPGPVPLLSMSVKSESLDDIGGHEKKSVTGSEVGGLDAQLWHACAGGMVQLPHVGAK
VVYFPQGHGEQAASTPEFPRTLVPNGSVPCRVVSVNFLADTETDEVFARICLQPEIGSSA
QDLTDDSLASPPLEKPASFAKTLTQSDANNGGGFSIPRYCAETIFPPLDYCIDPPVQTVLA
KDVHGEVWKFRHIYRGTPRRHLLTTGWSTFVNQKKLVAGDAIVFLRIASGELCVGVRR
SMRGVSNGESSSWHSSISNASTIRPSRWEVKGTESFSDFLGGVGDNGYALNSSIRSENQG
SPTTSSFARDRARVTAKSVLEAAALAVSGERFEVVYYPRASTAEFCVKAGLVKRALEQS
WYAGMRFKMAFETEDSSRISWFMGTIAAVQAADPVLWPSSPWRVLQVTWDEPDLLQG
VNRVSPWQLELVATLPMQLPPVSLPKKKLRTVQPQELPLQPPGLLSLPLAGTSNFGGHL
ATPWGSSVLLDDASVGMQGARHDQFNGLPTVDFRNSNYKHPREFSRDNQYQIQDHQV
FHPRPVLNEPPATNTGNYFSLLPSLQRRPDISPSIQPLAFMSASGSSQLETSSTKTAATSFF
LFGQFIDPSCTSKPQQRSTVINNASVAGDGKHPGTNNSSSDNKSEDKDNCRDVQPILNGI
AVRSGFRADIAAKKFQQSDSAHPTEASRGSQVSSLPWWQTQDAHKDQEFHGDSQTPHT
PASGSQ*

FIGURE 3F

Deduced amino acid sequence of MYB-1 from *Physcomitrella patens* (SEQ ID NO:22)

MKELNEDMEIPLGRDGEGMQSKQCPRGHWRPAEDDKLRELVSQFGPQNWNLIAEKLQ

GRSGKSCRLRWFNQLDPRINRHPFSEEEEERLLIAHKRYGNKWALIARLFPGRTDNAVK

NHWHVVTARQSRERTRTYGRIKGPVHRRGKGNRINTSALGNYHHDSKGALTAWIESKY

ATVEQSAEGLARSPCTGRGSPPLPTGFSIPQISGGAFHRPTNMSTSPLSDVTIESPKFSNSE

NAQIITAPVLQKPMGDPRSVCLPNSTVSDKQQVLQSNSIDGQISSGLQTSAIVAHDEKSG

VISMNHQAPDMSCVGLKSNFQGSLHPGAVRSSWNQSLPHCFGHSNKLVEECRSSTGAC

TERSEILQEQHSSLQFKCSTAYNTGRYQHENLCGPAFSQQDTANEVANFSTLAFSGLVK

HRQERLCKDSGSALKLGLSWVTSDSTLDLSVAKMSASQPEQSAPVAFIDFLGVGAA*

FIGURE 3G

Deduced amino acid sequence of CABF-3 from *Physcomitrella patens* (SEQ ID NO:23)

MADSYGHNAGSPESSPHSDNESGGHYRDQDASVREQDRFLPIANVSRIMKKALPSNAKI

SKDAKETVQECVSEFISFITGEASDKCQREKRKTINGDDLLWAMSTLGFEDYVEPLKVY

LHKYRELEGEKASTAKGGDQQGGKEGSQGVMGSMGMSGGMNGMNGTMNGNMHGH

GIPVSMQMLQQSYGQQAPPGMMYSPHQMMPQYQMPMQSGGNQPRGV

FIGURE 3H

Deduced amino acid sequence of SFL-1 from *Physcomitrella patens* (SEQ ID NO:24)

MMEAEQSYVQKLESLLGGVSTLVREEEETASVSEDEDDSNSLPQIQVAVKSKRKGERRK

RRERALERAEKVATDLASAPPLPKPKKPQLAADPSDPVRAYLRDIGRTKLLTAREEVDL

SHQIQDLLKLENIKSNLEREIGRNATIGEWSRAVGMEQNAFEARLKKGRFAKDKMVNSN

LRLVVSIAKNYQGRGMTLQDLIQEGSMGLVRGAEKFDPTKGFKFSTYAHWWIRQAVTR

SIADQSRTFRLPIHLYEVISRINKAKRMLVQEHGREARNEEVAELVGLTVEKLKSVVKSA

KAPGSMERPIGKDGDTTLGELVADTDVDSPEDAIVKQLMRQDIEGVLRTLNPREREVLR

LRFGLDDGRSKTLEEIGQIFKATRERIRQIEAKAMRKLRQPSRNSILREYLDVKSDAI*

FIGURE 4

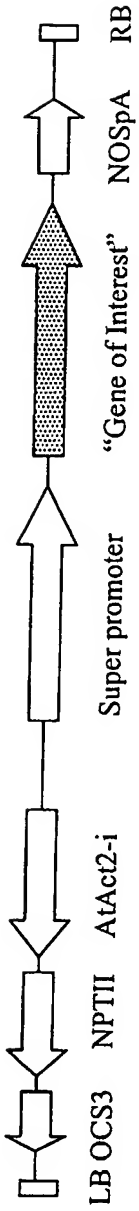

LB OCS3
NPTII
AtAct2-i
Super promoter
"Gene of Interest"
NOSpA
RB

Figure 5

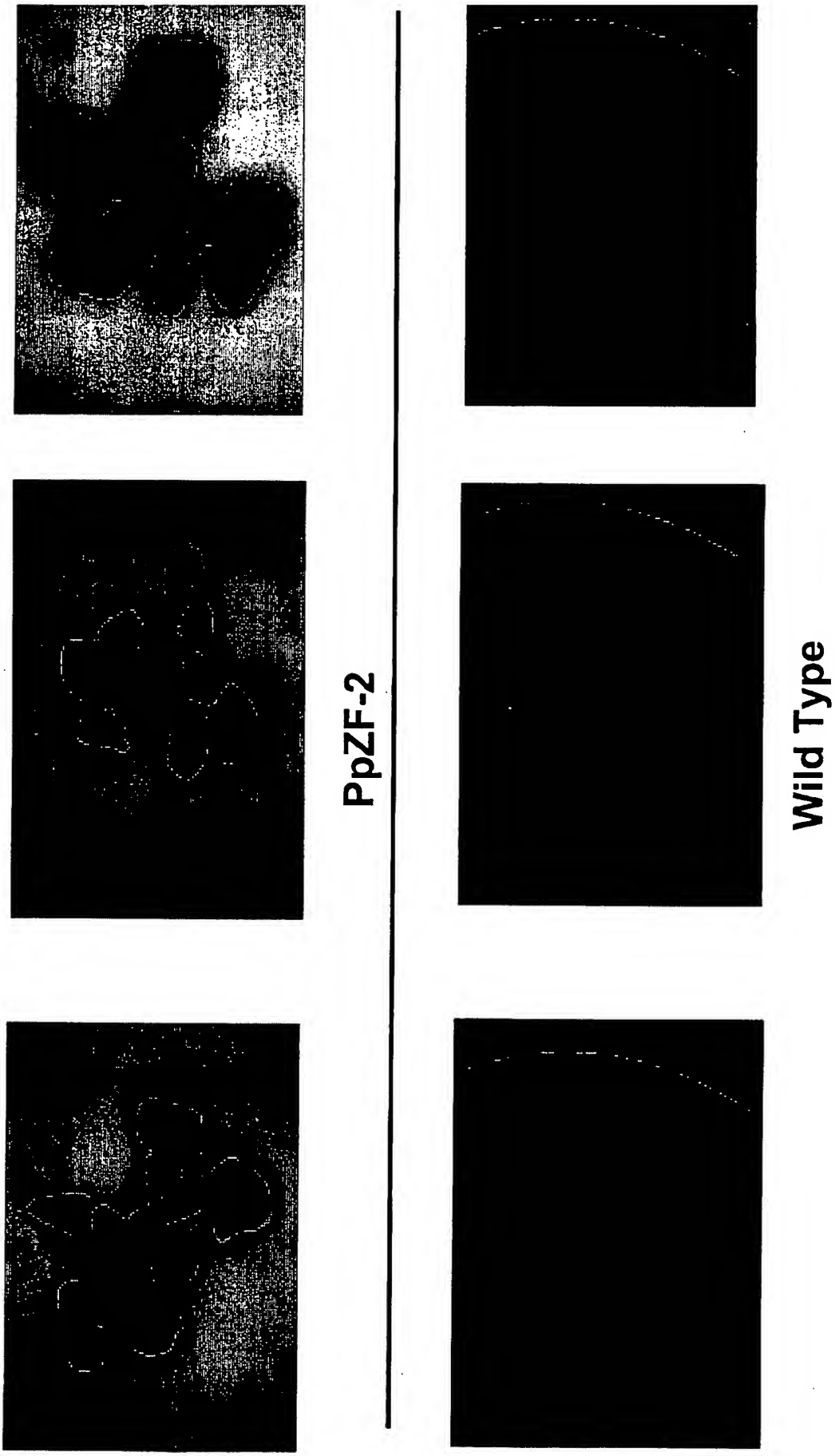


Figure 6

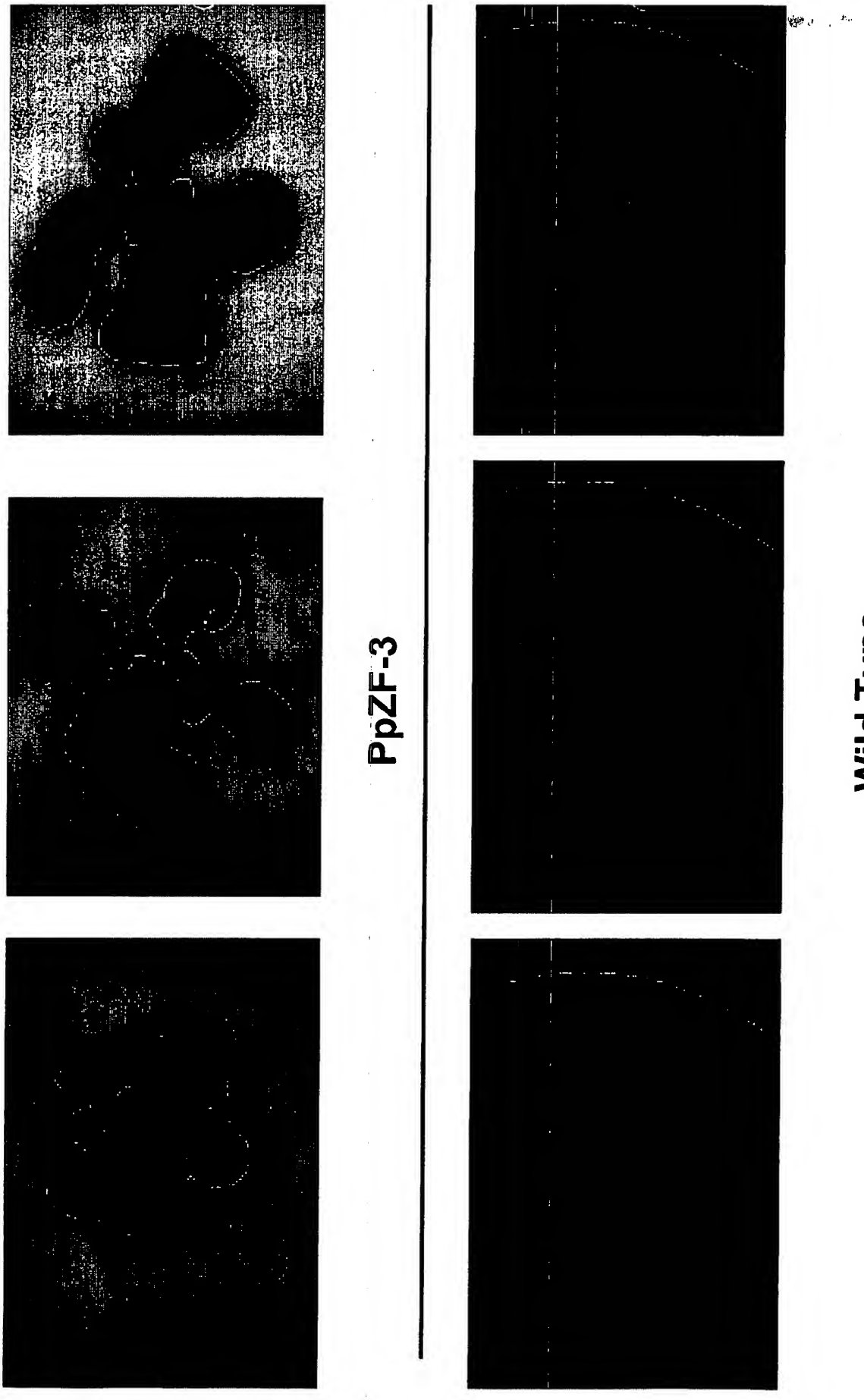


Figure 7

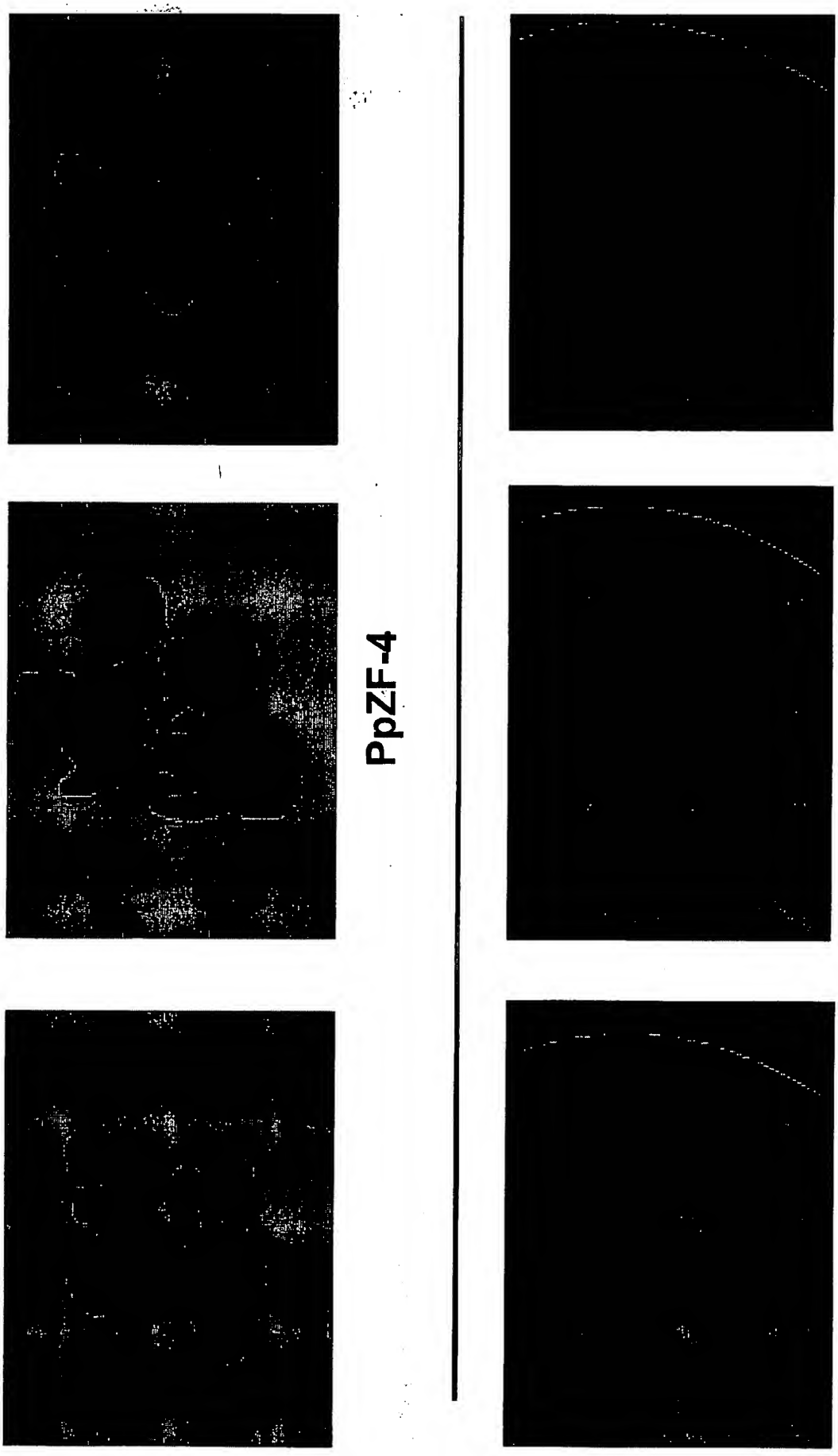


Figure 8

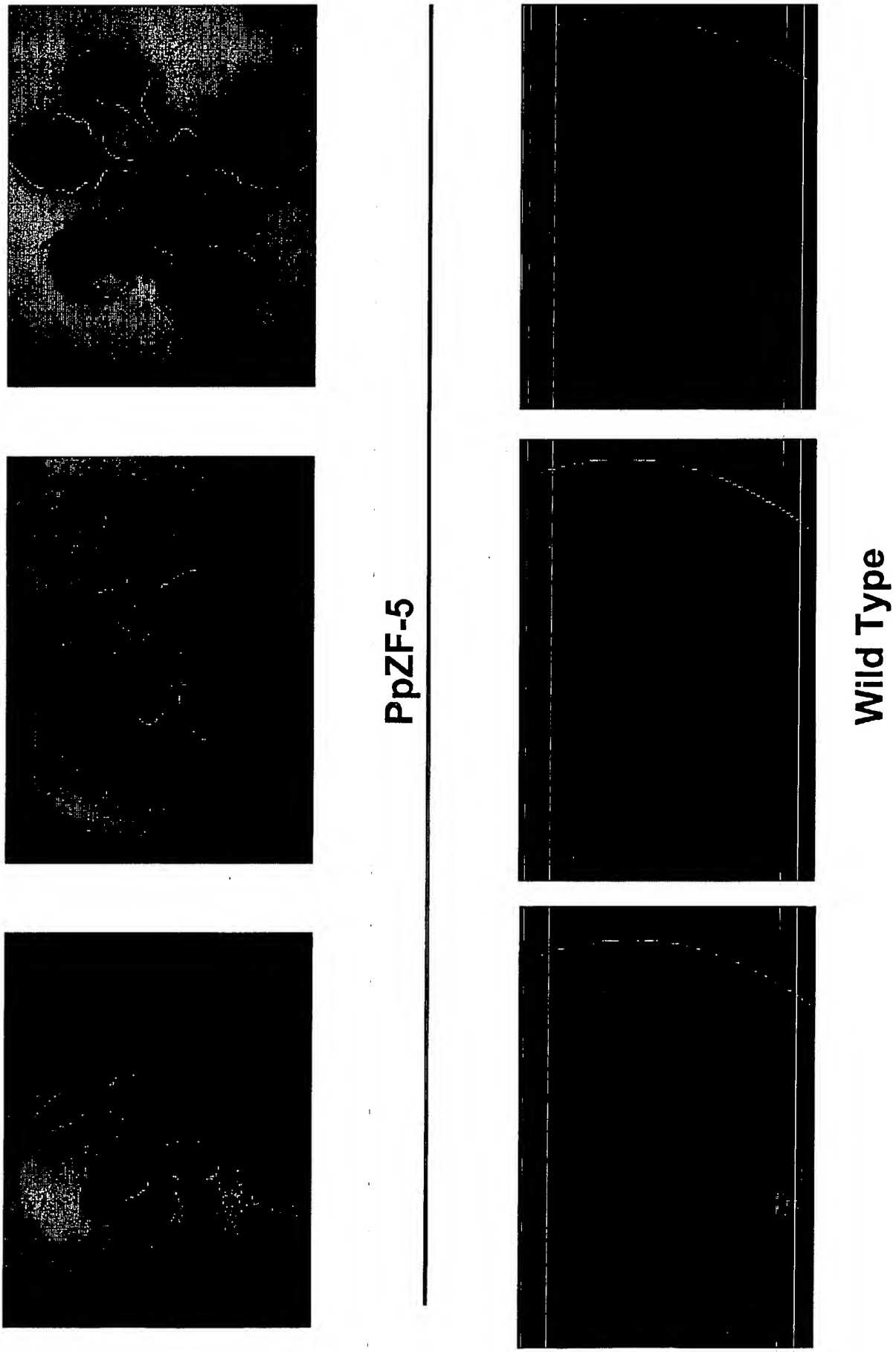


Figure 9

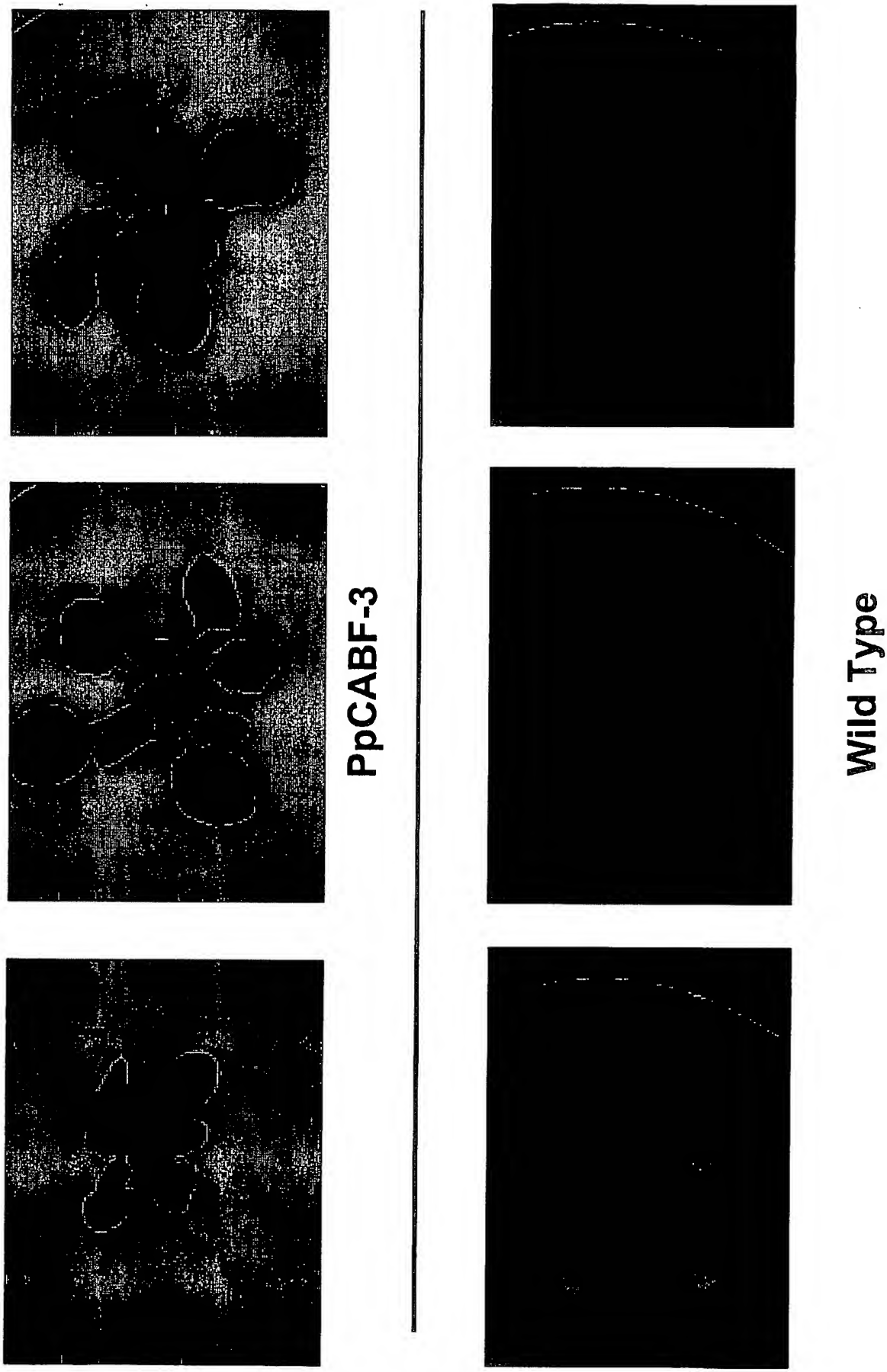


Figure 10

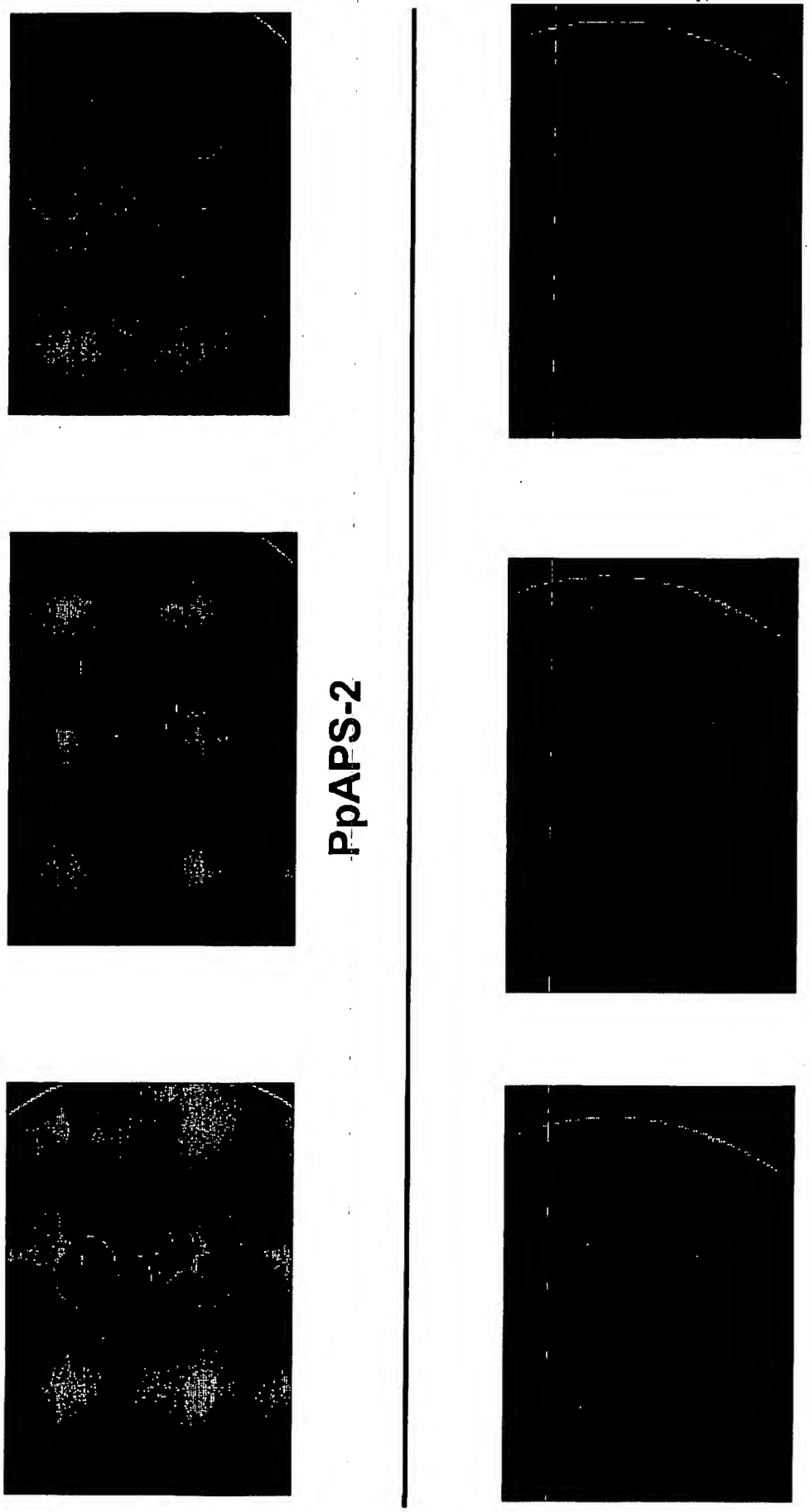


Figure 11

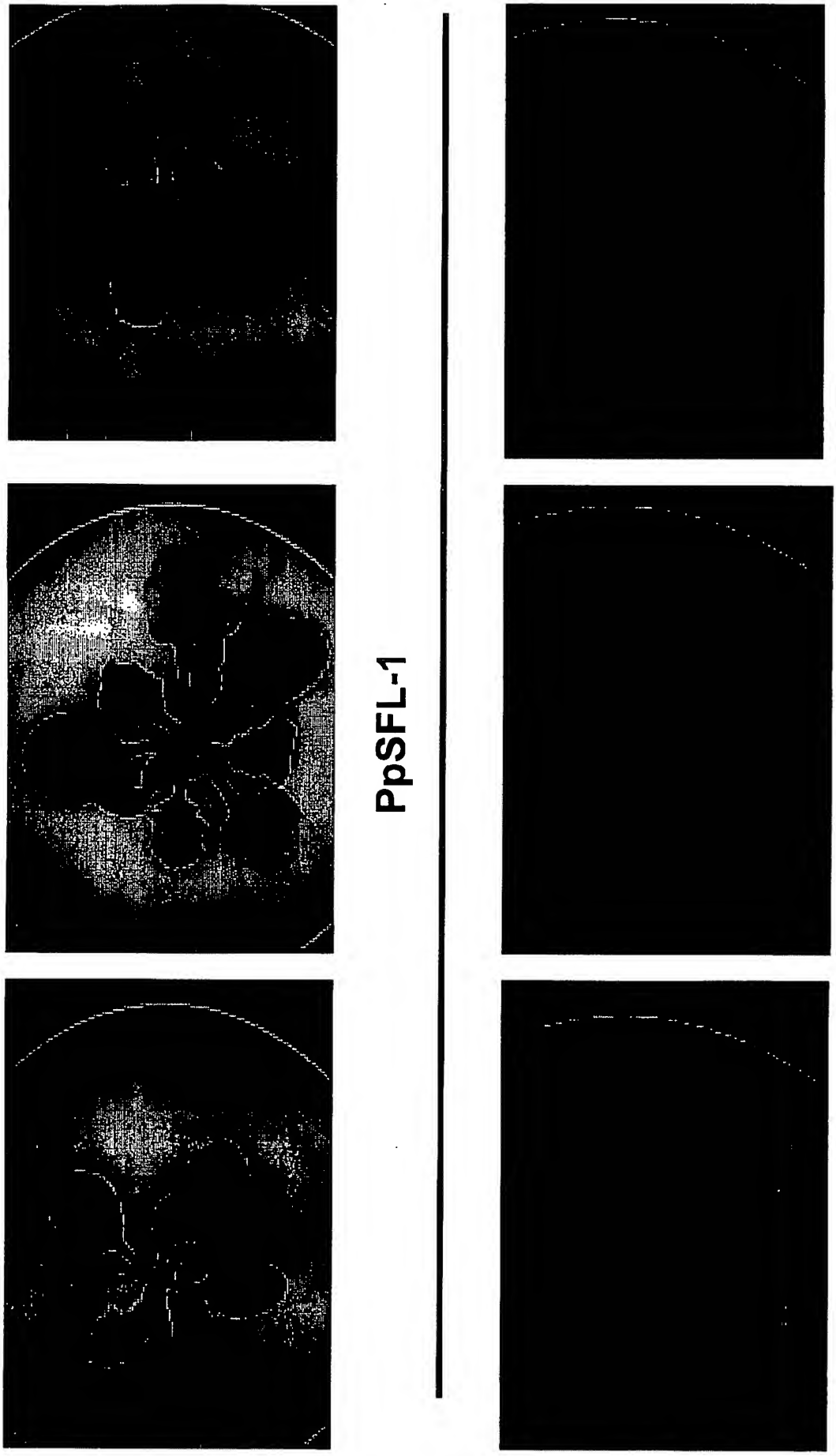


Figure 12

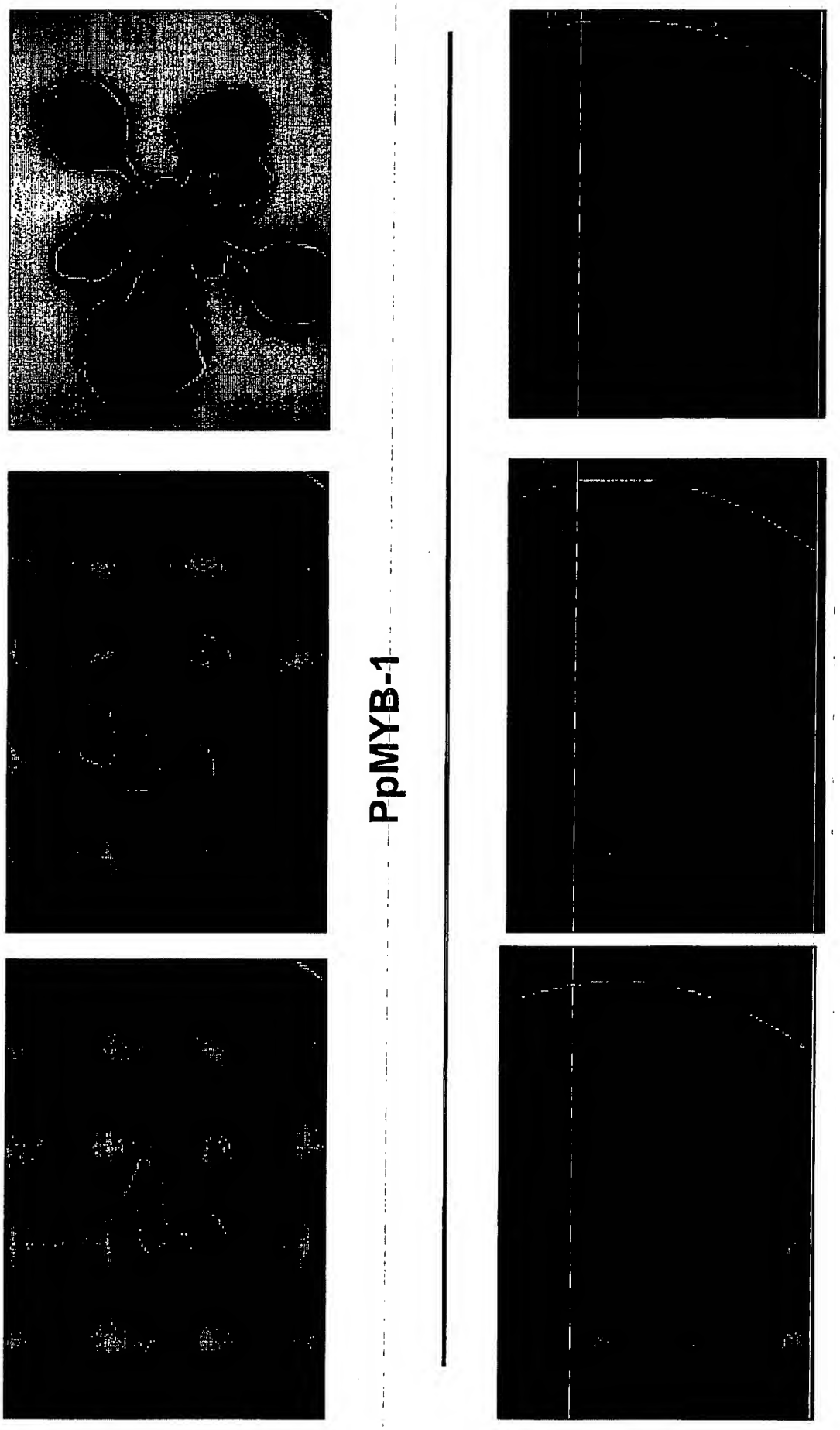


Figure 13

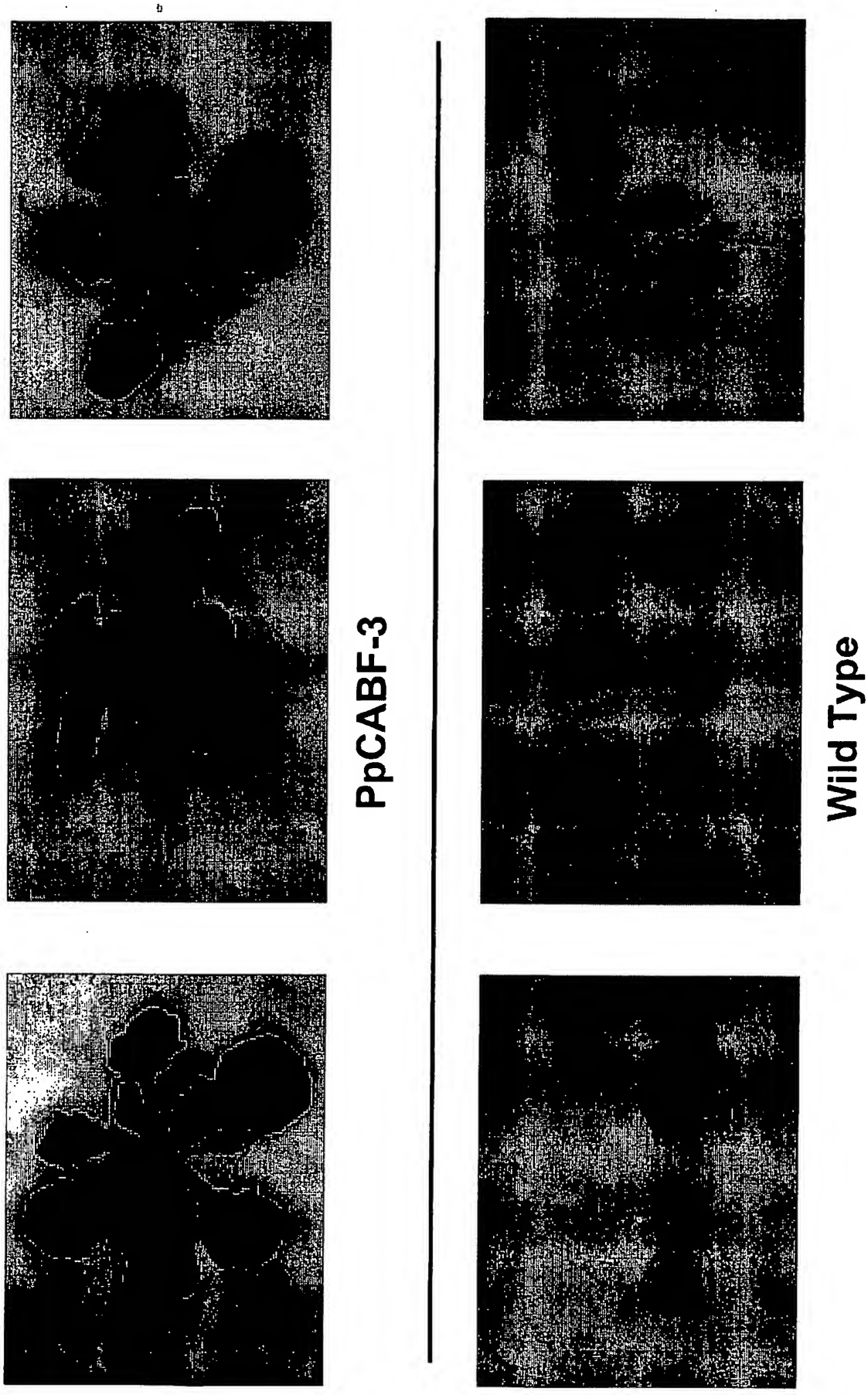


Figure 14

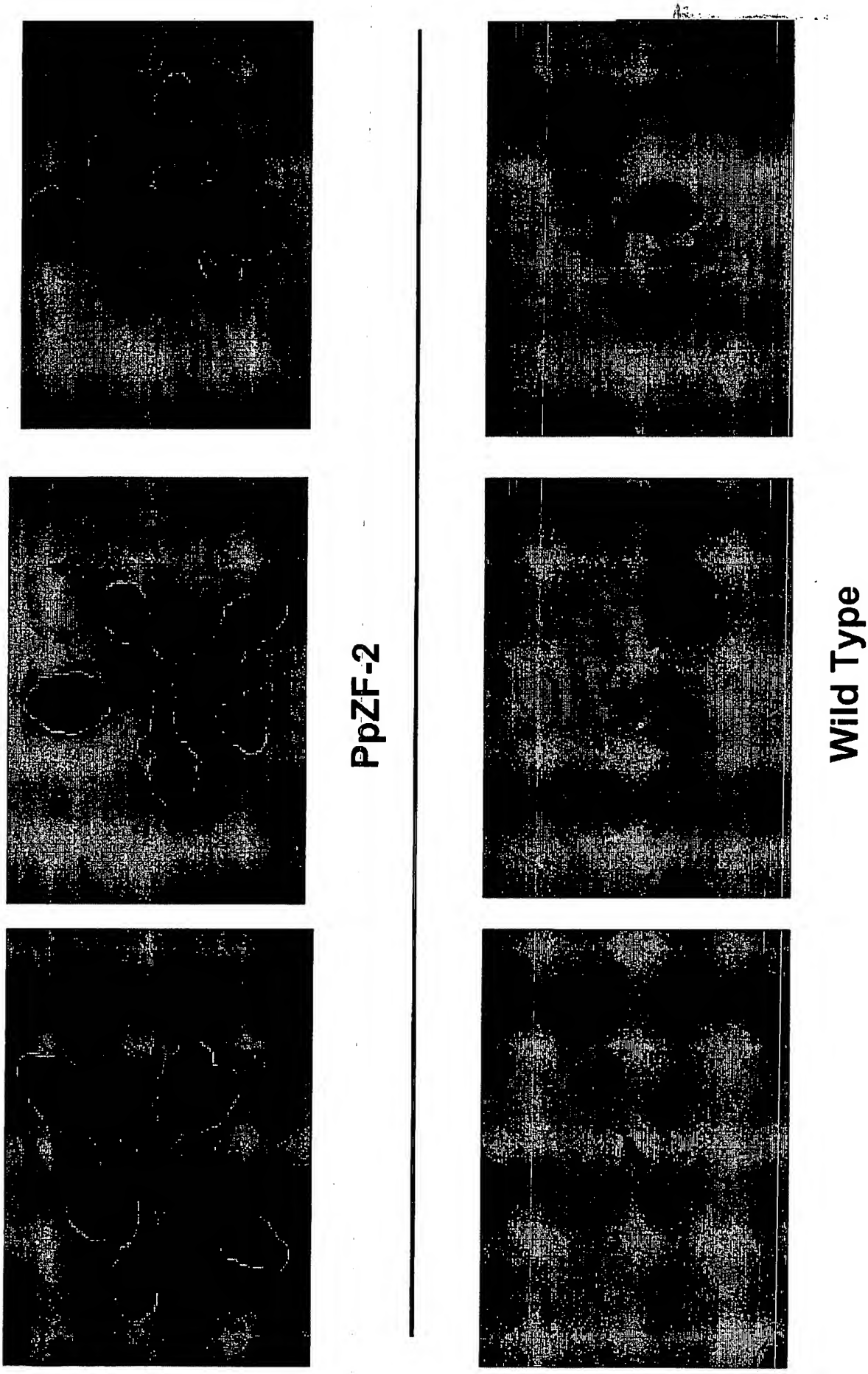
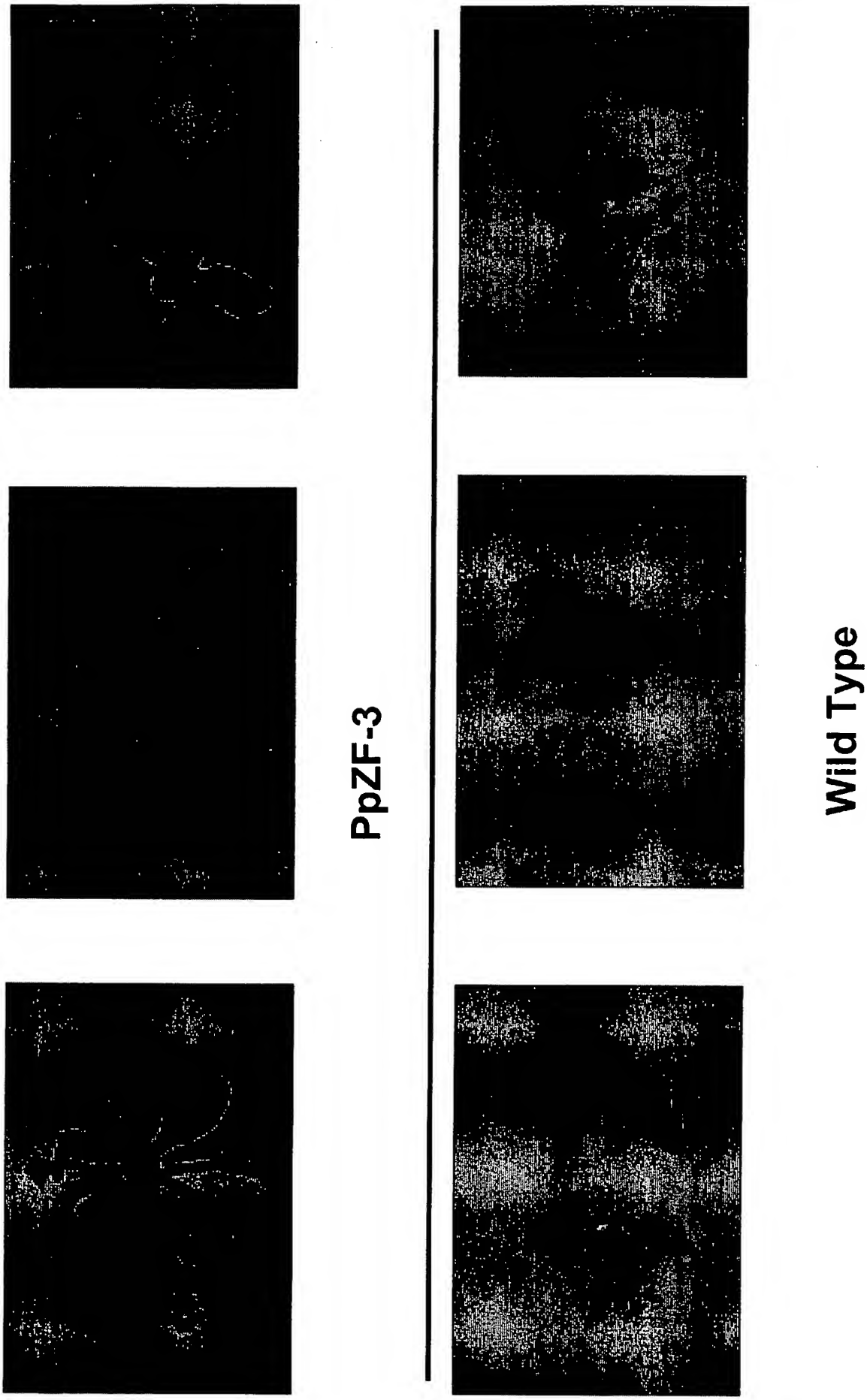


Figure 15



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 September 2001 (13.09.2001)

PCT

(10) International Publication Number
WO 01/66777 A1

(51) International Patent Classification[7]: C12N 15/82, A01H 5/00

(21) International Application Number: PCT/SE01/00451

(22) International Filing Date: 2 March 2001 (02.03.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

0000751-8	7 March 2000 (07.03.2000)	SE
60/240,319	13 October 2000 (13.10.2000)	US

(71) Applicant *(for all designated States except US)*: **SWETREE GENOMICS AB** [SE/SE]; P.O. Box 7984, S-907 19 Umeå (SE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ERIKSSON, Maria** [SE/SE]; Tvistevägen 46, S-907 36 Umeå (SE). **MORITZ, Thomas** [SE/SE]; Smörbäcksvägen 8, S-905 92 Umeå (SE). **ISRAELSSON, Maria** [SE/SE]; Rullstensgatan 27 B, S-906 55 Umeå (SE). **OLSSON, Olof** [SE/SE]; Långesandsvägen 33, S-430 90 Göteborg (SE).

(74) Agents: **HOLMBERG, Martin** et al.; Bergenstråhle & Lindvall AB, P.O. Box 17704, S-118 93 Stockholm (SE).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/66777 A1

(54) Title: TRANSGENIC TREES EXHIBITING INCREASED GROWTH, BIOMASS PRODUCTION AND XYLEM FIBRE LENGTH, AND METHODS FOR THEIR PRODUCTION

(57) Abstract: Important aims in nearly all tree-breeding programs around the world are to produce plants with increased growth rates and stem volumes, and shorter rotation times. Such trees would yield more biomass per area unit. Here the present inventors have shown that when over-expressing a key regulatory gene in the biosynthesis of the plant hormone gibberellin (GA) in hybrid aspen (*Populus tremula x P. tremuloides*), improvements in valuable traits such as growth rate and biomass are obtained. In addition, these trees also have longer xylem fibers than unmodified wild type plants. Long fibres are very desirable in the production of strong paper, but it has not (as yet) proved possible to influence this trait by traditional breeding techniques. A further advantage of the present invention is that it may reduce or eliminate the use of growth influencing chemicals in forestry.

Transgenic trees exhibiting increased growth, biomass production and xylem fibre length, and methods for their production

The present invention concerns transgenic trees exhibiting improved properties, and in particular properties of economic importance, such as increased growth rate, biomass production and xylem fibre length; transgenic plants, seeds, plant cells and other types of propagating material, as well as methods for their production.

Background

A major disadvantage with traditional tree breeding, especially for forest tree species, is the slow progress due to their long generation periods. However, by taking advantage of recent developments in gene technology the time required to produce a new variety could be reduced significantly. In addition, a biotechnological approach would allow closer targeting of traits considered desirable by the forest and pulp industries, in specific tree species.

To date, most applications of genetic engineering of trees have focused on modifying lignin biosynthesis, resulting in trees with less lignin or a modified lignin composition, earlier flowering, pest or herbicide resistance. In order to change growth and development processes in trees, the manipulation of plant hormone levels, or the hormone sensitivity, would also be of interest. However, as yet there has only been few examples of the modification of plant hormone levels in trees. These have mainly been accomplished by directly altering endogenous IAA biosynthesis or cytokinin biosynthesis or indirectly by modifying various hormone pools using the *Agrobacterium rolC* gene. Although such modifications in all cases lead to trees with altered growth characteristics and wood properties, so far no improvements with a clear practical application have been obtained.

Gibberellins (GAs) are a group of more than 100 tetracyclic diterpenes, some of which are essential endogenous regulators that influence growth and development processes throughout the plant life cycle, e.g. shoot elongation, the expansion and shape of leaves, flowering and seed germination. The best examples illustrating the importance of GAs in control of shoot elongation are GA-deficient mutants of *Arabidopsis*, maize and pea. These have reduced levels of active GA(s) compared to wild type plants,

resulting in a dwarfed phenotype due to a reduction in internode length. The phenotype of such GA-deficient mutants can be completely restored by the application of an active GA. At the cellular level, GAs have been found to promote both cell division and cell elongation.

Biosynthesis of GAs *in planta* occurs through the isoprenoid pathway from mevalonic acid. Gibberellin levels are mainly regulated by transcriptional control of gibberellin biosynthesis genes. In particular, the multifunctional enzyme gibberellin 20-oxidase (GA 20-ox) is a key-enzyme in controlling GA biosynthesis (Fig. 1). It catalyses the stepwise conversion of the C-20 gibberellins, GA_{12}/GA_{53}, by three successive oxidations to GA_9/GA_{20}, which are the immediate precursors of the active gibberellins, GA_4 and GA_1, respectively. The expression of the GA 20-oxidase gene is down regulated by the action of $GA_{1,4}$, suggesting that direct end-product repression is involved in regulation of the gene. In addition, some authors have suggested that GA 20-oxidase is photoperiodically regulated at the transcription level.

Application of chemicals that alter GA levels in the plant is a common practise in traditional agriculture and horticulture. Inhibitors of GA biosynthesis are especially commonly used as growth retardants in cereals and ornamental plants. In order to reduce the use of these chemicals, a biological approach like genetic modification of endogenous GA biosynthesis would have clear advantages. Using *Arabidopsis* as a model organism it has been shown that it is possible to change GA levels by modifying GA 20-oxidase enzyme levels and that this results in plants with altered growth and development patterns. Transgenic *Arabidopsis* expressing the GA 20-oxidase in a sense orientation shows earlier flowering and taller stems than wild type plants, whereas antisense plants have the reverse properties [Coles, J.P. *et al.* Modification of gibberellin production and plant development in Arabidopsis by sense and antisense expression of gibberellin 20-oxidase genes. *Plant J.* **17**, 547-556 (1999)].

Modification of GA biosynthesis in a higher species, such as trees would be of additional interest since this would open up ways to modify wood. Previous hormone application studies have shown that GAs are required for the differentiation of xylem fibres, and that they have pronounced effects on the length of secondary xylem fibres and on both longitudinal and radial growth in hard wood species and conifers.

Obviously there remains a need for improved methods for modification of the growth properties of trees, in particular properties of technical and economical interest, such as growth rate, biomass increase and fibre length. Likewise, there remains a need of transgenic trees, exhibiting improved properties, such as increased growth rate, stem volume and xylem fibre length. Consequently, the objective of the present invention is to provide such improved methods and transgenic trees. Another objective is to reduce or eliminate the use of growth influencing chemicals in forestry.

Prior art

It has been shown (Huang, S. *et al.*, Overexpression of 20-Oxidase confers a gibberellin-overproduction phenotype in Arabidopsis. Plant Physiol., 1998, vol. 118, p 773-781) that the level of active GAs were raised following overexpression of GA 20-oxidase in *Arabidopsis thaliana*. This finding is however not directly transferable on the present invention.

Most dicotyledons and all gymnosperm undergo some degree of secondary thickening. The amount of thickening depends on whether the mature plant is a herbaceous or a woody (arborescent) plant. For example an *Arabidopsis* plant will only produce secondary thickening under special conditions whereas a woody species, for example *Populus*, will undergo a high degree of secondary growth. The low degree of secondary thickening in herbaceous species will also make it very difficult to predict how specific genetic changes in herbaceous species corresponds in woody species in regards to changes in wood formation.

As an example, by over expressing the GA 20-oxidase in *Arabidopsis thaliana* the level of active GAs in the plant can be raised, as shown by Huang *et al.* (Supra). The transgenic plant phenotype included a dramatic cell elongation in all tissues. The petioles, inflorescence stems and leaves all showed cell elongation. It is clear from these results that the enhanced gibberellin levels have an effect on cell elongation from germination. However, from these results it is impossible to predict what the effect of GA 20-oxidase over expression in a woody plant would be. Cell division in the cambial mersistem in a woody plant is under the control of several different hormones and physical constraints not found in an annual plant such as *Arabidopsis thaliana*.

It was therefore surprising that there would be increased cell elongation of cells (fibres) originating from the cambium, and that it would be an increase in biomass caused by increased cambial cell division. It would not be obvious to one skilled in the art to over express GA 20-oxidase to increase wood biomass and fibre elongation.

Increase in the diameter of tree stems occurs primarily from meristematic activity in the vascular cambium, a cylindrical lateral meristem located between the xylem and phloem of the stem, branches, and woody roots. Two types of cell division occur in the cambium: additive and multiplicative. Additive division involves periclinal division of fusiform cambial initials to produce xylem and phloem mother cells that in turn divide to produce xylem and phloem cells. Multiplicative division involves anticlinal division of fusiform initials that provide for circumferential expansion of the cambium. After xylem and phloem cells are cut off by the cambial mother cells, they differentiate in an ordered sequence of events that includes cell enlargement, secondary wall formation, lignification and loss of protoplasts. These events do not occur stepwise, but rather as overlapping phases.

Summary of the invention

It is an object of the present invention to provide transgenic plants, seeds and plant cells, exhibiting improved growth parameters and in particular improved parameters of economical interest, such as increased growth rate, biomass increase and xylem fibre length. It is another object of the invention to provide a method for the production and proliferation of said transgenic plants. The above objects and others not explicitly mentioned, are fulfilled through a transgenic woody plant having a DNA sequence coding for the expression of a polypeptide exhibiting GA 20-oxidase activity functionally inserted in the plant genome. Further embodiments are as defined in the attached dependent and independent claims, incorporated herein by reference.

Other features and advantages of the invention will be apparent form the following, non-limiting description and examples, and from the claims.

Brief description of the drawings

The invention will be described in closer detail below, in the description and accompanying examples and drawings, in which

Figure 1 shows the biosynthetic pathways converting GA_{12} and GA_{53} to the biologically active products GA_4 and GA_1, and their deactivated catabolites, GA_{34} and GA_8. GA 20-oxidase catalyses the oxidation at the C-20 carbon.

Figure 2 shows a Northern analysis of ten GA 20-oxidase over-expressing lines (numbers 1 to 15) and the non-transformed control (C). 18µg of total RNA was loaded from each sample, and probed with a labelled fragment isolated from *Arabidopsis* gibberellin 20-oxidase (*AtGA20ox1*) cDNA or from an endogenous ubiquitin (*UBQ*) EST sequence (*ptUBQ*).

Figure 3 shows the enhanced growth of transgenic hybrid aspen: the cumulative shoot elongation (upper diagram) and diameter growth (lower diagram) of various transgenic GA 20-oxidase over-expressing lines, after generation from tissue culture, potting and cultivation in a growth chamber for seven weeks (at time zero).

Figure 4 is a photograph showing control and transgenic (line 11) plants after 12 weeks in the growth chamber.

Figure 5 illustrates the effects of GA 20-oxidase over-expression on (A) Cell lengths, (B) cell numbers per internode, (C) number of xylem fibres and (D) xylem fibre length. Cell lengths and cell numbers were measured in fully elongated internodes of actively growing plants as was number of xylem fibres. Data for the GA 20-oxidase over-expressing lines are means for nine independently generated lines, respectively. Fibre length values for non-elongating plants represent means for three independently generated lines. The vertical bars represent SE.

Figure 6 shows that there is enhanced diameter growth of transgenic hybrid aspen also in the second growth season. The cumulative diameter growth of control (WT) and line 4 and 13 was determined under a period of 2,5 month, beginning after breaking of dormancy.

Description of the invention

The present inventors have surprisingly shown that ectopic over-expression of a GA 20-oxidase gene in trees, here exemplified by a temperate-zone deciduous tree, results in significant changes in growth rate, stem volume and xylem fibre length. A cDNA sequence (SEQ. ID. NO. 1) homologous to the *Arabidopsis thaliana* GA 20 oxidase sequence X83379 (EMBL accession number) encoding a polypetide exhibiting GA 20 oxidase activity was used. Thus, the results obtained by the present inventors show that genetic modification of GA levels in trees can be used to modify traits that are extremely important for the forest, pulp and paper industries.

The results show that an endogenous elevation of the biologically active gibberellins GA_1 and GA_4 in trees will accelerate growth. This was demonstrated by expressing the *AtGA20ox1* gene from *Arabidopsis* in hybrid aspen, *Populus tremula x tremuloides*, resulting in trees with faster height and diameter growth, larger leaves, longer xylem fibres and increased biomass. This phenotype is partly reminiscent of traits previously observed in transgenic *Arabidopsis* plants over-expressing the same GA 20-oxidase (Coles, J.P. *et al.*, Modification of gibberellin production and plant development in Arabidopsis by sense and antisense expression of gibberellin 20-oxidase genes. *Plant J.* **17**, 547-556 (1999). Such plants had longer hypocotyls, larger rosette leaves, longer petioles, and accelerated flowering compared to wild type plants.

However, studies regarding biomass increase and anatomical changes in transgenic *Arabidopsis* have not been reported. It is highly doubtful if an increased biomass would be achieved in transgenic *Arabidopsis*.

There have been suggestions that the GA 20-oxidase is a key enzyme in the regulation of GA controlled growth [Hedden, P. & Kamiya, Y. Gibberellin biosynthesis: Enzymes, genes and their regulation. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**, 431-460 (1997)]. The fact that transcription of the GA 20-oxidase gene is feedback regulated by the action of GAs, suggests that constitutive over-expression of the GA 20-oxidase gene will disturb the endogenous regulation of GA homeostasis. Here, over-expression of GA 20-oxidase in hybrid aspen resulted in an increase in C_{19}-GAs involved in both the early 13-hydroxylation pathway and the non-13-hydroxylation

pathway (Fig. 1). Thus, there was no consistent difference in levels between GA_1 and GA_4, both of which are biologically active in *Populus*. The present results therefore are in agreement with previous GA measurements on transgenic GA 20-oxidase overproducing *Arabidopsis* plants. Furthermore, in both hybrid aspen and *Arabidopsis* there was a dramatic increase in the levels of the deactivated end-product, GA_8 (Table 1).

This has been suggested to be due to a slower turnover of GA_8 than GA_1. The relatively low levels of the other end-product, GA_{34}, formed by the non-13-hydroxylation pathway can be explained by a more rapid turnover of this compound than GA_8. In contrast to the C_{19}-GAs, a decrease in levels of C_{20}-GAs was observed in the transgenic plants. This is consistent with the fact that C_{20}-GAs are, by definition, substrates for the GA 20-oxidase (Table 1).

It is surprising that the elevated GA 20-oxidase expression has such marked effects in *Populus tremula x tremuloides*. This suggests that constitutive over-expression of the GA 20-oxidase gene disturb the endogenous regulation of GA homeostasis.

In the present study the inventors found differences in GA levels between leaf and stem tissues: increases in levels of the active GA_1 and GA_4 being highest in stem tissue. This could have been due to the *AtGA20ox1* gene being expressed most strongly in the stem. In hybrid aspen, however, the CaMV 35S promoter is generally slightly more active in leaves than in the stem, although exceptions are known to occur. Therefore, the higher GA_1 and GA_4 levels in the stem are more likely due to differences in transport of GAs or in the biosynthesis of GA precursors. However, it can not be ruled out that there may be a difference in GA 3ß-hydroxylase or GA 2ß-hydroxylase activity in the leaves. In *Arabidopsis* it has been shown that 2ß-hydroxylation is activated at the transcriptional level by active GAs.

The observed increase in height growth (Figure 3) in the transgenic GA 20-oxidase *Populus* trees was apparently not a consequence of increased cell elongation, as no significant changes in cell lengths were found (Figure 5A). However, observed differences in numbers of cells per fully elongated internode clearly indicate that elevated GA levels affect cell divisions in stem tissue (Figure 5). This is an interesting

aspect of the present invention, which opens up many possibilities of influencing the production of fibrous raw material and biomass for different purposes.

Although GAs have been found to affect cell elongation, it is also known GAs can induce mitotic activity in the subapical region of the stem. It must be emphasised, however, that although samples were taken for cell length measurements from the same position for both control and transgenic plants (as defined by the number of internodes from the apex), it can not be ruled out that the cells in the transgenic plants could have been at an earlier developmental stage than in wild type plants, and they could still have been elongating. It has also been suggested that the action of GAs in meristematic tissues extends the elongation zone of the organ being formed.

Table 1. Concentration (ng g^{-1} fresh weight) of GAs in stem/leaf tissue from internodes 7 and 8 (see text for details) of transgenic hybrid aspen expressing the *AtGA20ox1* gene. Means of three independent measurements. n.d. = not detectable.

Genotype	GA_{12}	GA_{53}	GA_{19}	GA_9	GA_{20}	GA_4	GA_1	GA_{11}	GA_8
Control	5.78/1.15	3.22/0.41	3.15/1.21	n.d./n.d.	0.45/0.43	0.84/0.88	0.63/0.63	1.63/1.23	15.0/9.4
Line 1	1.16/0.79	0.29/0.28	0.85/0.55	3.20/0.64	5.36/1.44	9.71/4.03	20.2/1.29	5.00/3.78	58.6/40.7
Line 2	3.02/0.90	1.04/1.25	3.55/0.59	4.34/1.24	3.69/2.84	9.04/6.60	10.2/2.47	8.37/5.94	77.4/43.8
Line 4	1.06/0.71	0.27/0.24	0.81/0.63	5.23/0.86	5.03/2.84	11.3/4.76	11.7/1.58	7.84/4.78	53.2/52.9
Line 7	1.00/0.73	0.29/0.28	1.21/0.60	5.38/0.78	4.74/2.36	20.7/6.29	14.0/1.35	6.03/4.69	68.4/32.7
Line 9	2.49/0.73	0.62/0.26	1.19/0.56	5.67/0.73	6.00/2.16	9.70/5.13	5.42/2.35	10.46/4.85	70.0/49.6
Line 11	1.67/0.62	0.57/0.28	2.04/0.60	4.31/1.42	5.77/3.31	10.7/5.30	14.2/1.81	9.43/4.28	72.6/42.6
Line 13	0.91/0.78	0.26/0.25	0.90/0.64	6.58/0.64	16.4/3.45	19.8/6.46	29.3/3.22	7.98/3.85	75.9/55.1
Line 14	1.26/0.64	0.27/0.25	0.62/0.55	3.2/0.41	4.01/1.48	6.99/2.36	11.6/1.38	5.52/4.69	78.6/37.5
Line 15	1.82/0.60	-	5.57/0.74	-	2.53/2.57	3.14/2.58	2.58/1.62	7.53/5.02	57.6/48.2

Studies involving the application of plant hormones have shown that GAs can increase cambial activity in trees, especially in conjunction with IAA. Here, the engineered increase in GA levels resulted in faster stem diameter growth (Fig. 4). An effect of elevated GA levels on secondary growth was also observed with the increase in number of xylem fibre cells and xylem fibre length (Fig. 5C, D). The length increase measured in the experiments is about 8 % which must be held to be a significant increase, having consequences for the fibre properties and thus the potential uses for the fibrous raw material.

This is consistent with application studies with both GAs and inhibitors of GA biosynthesis. In *Populus*, fibre and vessel lengths increase in the transition zone between juvenile and mature wood. Therefore, the time it takes for a cell to mature within the cambial region, i.e. the time spent in the different differentiation zones, will ultimately determine its size and cell wall thickness. If the action of GAs like GA_1 and/or GA_4 extends this transition time, this would explain both the increased fibre length in the mature parts of the transgenic plant and the lack of differences in xylem fibre length between transgenic and control plants in young stem tissue.

The increase in leaf size in the transgenic plants was a result of an increase in both radial and longitudinal elongation (Table 2). At earlier stages the longitudinal growth was more pronounced, resulting in long, narrow leaves. However, at later stages the leaf morphology became similar to control plants with a similar leaf width to length ratio. It has been shown that GAs promote leaf elongation e.g. in garden pea. The very long and narrow leaves observed in constitutive GA signal mutants, like *sln* mutants of barley and *spy* mutants of *Arabidopsis*, also indicate that GAs promote leaf elongation. However, little is still known on the more specific regulatory roles of GAs in leaf development. In *Arabidopsis*, leaf blade expansion is regulated by at least two independent and polarised processes: length and width development, with *ROTUNDIFOLIA* and *ANGUSTIFOLIA* genes playing specific roles. It is not known whether GAs modulate the cellular responses to these genes. However, it is noteworthy that leaf phenotypes in the transgenic GA 20-oxidase aspen change during development, suggesting that the two developmentally separate processes of leaf length and width growth are both affected by GAs, but at different stages.

Table 2. Morphological characterisation of wild type and transgenic hybrid aspen expressing the *AtGA20ox1 gene*. The numbers of plants used for the measurements were 10, 5 and 6 for the control, line 2, and line 11, respectively. In the biomass determination the numbers of plants sampled were 7, 2, 3 and 23 for the control, line 2, line 11, and all sampled transgenics, pooled, respectively.

For all statistical analyses ANOVA was used to compare the plants with respect to genotype. Statistically significant difference are indicated at the 1% (*) and 5 % (**) probability levels (Fisher's PLSD).

	Genotype		
	Control	Line 2	Line 11
Internode length (cm ±SE)	2.19 ±0.05	3.06 ±0.09*	3.07 ±0.09*
Leaf length and width (cm ±SE)			
Leaf length 10	8.34 ±0.36	9.52 ±0.38**	9.17 ±0.28
Leaf width 10	7.25 ±0.41	6.28 ±0.20	5.85 ±0.24**
Leaf length 14	9.52 ±0.45	11.18 ±0.60**	10.78 ±0.41
Leaf width 14	8.00 ±0.38	8.35 ±0.39	8.52 ±0.25
Leaf length 18	9.90 ±0.79	11.06 ±1.00	11.68 ±0.56
Leaf width 18	7.48 ±0.64	8.80 ±0.47	8.88 ±0.43
Leaf length 20	7.75 ±0.34	11.94 ±0.44*	11.25 ±0.48*
Leaf width 20	6.07 ±0.46	9.44 ±0.41*	8.85 ±0.41*
Dry Weight Biomass (g)	Control	Line 2, 11	Transgenes
Leaf	5.45 ±0.51	6.38 ±0.58	6.14 ±0.28
Stem	4.58 ±0.43	10.42 ±1.01*	10.34 ±0.49*
Root	43.05 ±4.71	38.36 ±2.63	37.81 ±2.06
Fresh Weight Biomass (g)			
Leaf	17.10 ±1.63	23.10 ±2.01**	22.05 ±0.91**
Stem	12.04 ±1.19	31.39 ±2.88*	30.58 ±1.28*

Biomass measurements in transgenic and wild type plants revealed that an increase in GA 20-oxidase activity leads to a general increase in growth. This effect was especially pronounced in the stem, indicating that GAs have a strong effect on stem growth. Spraying experiments with various GAs have previously shown that these hormones tend to increase shoot biomass in *Populus* at the expense of root growth. Reductions in root formation due to increased GA levels in roots have earlier been shown in trees. In this study, poor root initiation was the major problem for survival of the transgenic GA 20-oxidase plants in tissue culture and when planting in soil. At later developmental stages, root development was still affected, but to a lesser degree (Table 2). It has also been shown that the effect of GAs on rooting varies with the stage of root development.

In conclusion, the present inventors have here demonstrated the important role of GA 20-oxidase activity in GA-controlled growth in *Populus*. It is surprising that the effects, hitherto only achieved, and in some cases only partially achieved, by external application, e.g. spraying, can be achieved by endogenous expression. It is also surprising that the inventive method can be applied to a higher species, here represented by hybrid aspen, *Populus tremula x P. tremuloides*. One important advantage of the present invention is thus that it may make it possible to reduce or eliminate the use of growth influencing chemicals in forestry.

The present invention opens up ways to genetically engineer trees to grow faster and produce more biomass simply by increasing endogenous GA levels. Interestingly, fibre lengths also increase as a result of the over-expression of the GAs. These results have both practical implications (assisting in the production of modified trees of interest to the pulp, paper and forest industries), and scientific importance, by allowing previously impossible studies on the mechanisms whereby GAs control growth and development in trees.

Examples

Experimental Protocol:

Plant Vector Construction

The *Arabidopsis GA 20-oxidase* cDNA construct cloned in bluescript pAt2301 [Phillips. A.L. *et al.* Isolation and expression of three gibberellin 20-oxidase cDNA clones from *Arabidopsis*. *Plant Physiol.* 108. 1049-1057 (1995)], recently renamed *AtGA20ox1* [Coles. J.P. *et al.* Modification of gibberellin production and plant development in Arabidopsis by sense and antisense expression of gibberellin 20-oxidase genes. *Plant J.* 17. 547-556 (1999)] was obtained as a gift from P. Hedden.

An upstream ATG sequence. preceding the translational start of the GA 20-oxidase enzyme. was removed by *in vitro* mutagenesis. A forward primer matching the 5'-end of the GA 20 oxidase gene. but lacking the extra ATG and carrying a *Bam HI* and *Xba I* site. and a reverse primer spanning an internal *Hind III* site in the GA 20 oxidase gene, were used in a PCR reaction. Using *AtGA20ox1* as a template. a PCR product was obtained and subcloned into the pOK12 vector, generating plasmid pOK12.GA20ox5'. An entire GA 20-oxidase cDNA lacking the extra ATG was then re-constituted. by digesting *AtGA20ox1*. isolating a fragment containing the 3'-end of the GA 20-oxidase gene. and ligation of this fragment into pOK12.GA20ox5'. giving plasmid pOK12.AtGA20ox1. The GA 20-oxidase cDNA was then isolated from pOK12.AtGA20ox1, by digestion with *Bam HI*. and ligated in sense orientation into the *Bam HI* site of the binary vector pPCV702.kana [Walden. R., Koncz. C. & Schell. J. Methods Mol. Cell. Biol. 1. 175-194 (1990). The use of gene vectors in plant molecular biology. *Methods Mol. Cell Biol.* 1. 175-194 (1990)], subjecting it to the control of the CaMV 35S promoter.

Plant Transformation and Growth Conditions

Hybrid aspen *Populus tremula* L. x *P. tremuloides* Michx. clone T89 was transformed and regenerated essentially as previously described [Nilsson. O. *et al.* Spatial pattern of cauliflower mosaic virus 35S promoter-luceferase expression in transgenic hybrid aspen trees monitored by enzymatic assay and non-destructive

imaging. *Transgenic Reaserch* 1. 209-220 (1992)]. Out of 14 independent lines. 10 were selected and multiplied by *in vitro* shoot culture on half-strength MS-medium containing minerals and vitamins [Murashige. T. & Skoog. F. A revised medium for rapid growth and bio-assay with tobacco tissue cultures. *Physiologia Plantarum* 15. 473-479 (1962)].

Following root initiation. the plants were dipped in Weibufix rooting powder (Svalöf Weibull AB. Hammenhög. Sweden). and potted in a fertilised peat: perlite mixture (5:1) and kept in a growth chamber at 18°C under a photoperiod of 18 h and a relative humidity of 90%. The photon flux density of the main light period (10 h) was 175 µmol m^{-2} s^{-1} at 400 –750 nm (Osram Power Staw HQI-TS 400 W/D lamps. Osram. Germany) and daylength extensions were given using low-intensity light (20 µmol m^{-2} s^{-1}). Plants were watered daily. and repotted and fertilised with a complete nutrient solution (SuperbaS. Supra Hydro AB. Landskrona. Sweden) when needed. After 108 days some plants were transferred to short photoperiod conditions (10 h) to induce growth cessation. These plants were then cold acclimated and kept dormant at 8°C for 4 weeks. at which point samples were taken for fibre length measurements.

Growth Measurements and Sampling

At seven weeks of age. the plants were marked at an actively growing internode at about the same position in all plants. This was used as a reference point for diameter growth measurements and for counting the internodes. The diameter. height and number of internodes of the plants were measured every 3rd to 4th day. The number of internodes was counted from the top to the reference point: the first internode being defined as the one below the uppermost leaf at least 1 cm long. Plants grown under long days were harvested after 100 days, using seven control plants and 25 plants representing nine transgenic individual lines (line 6 was excluded). Internodes 7 and 8. with the upper leaves included. were sampled for GA analysis and leaves from nodes 9 and 10 for northern analysis. All tissues were frozen in liquid N_2 immediately after sampling. For anatomical studies. the length of internodes 15 and 16 were measured. then excised and fixed immediately in FAA. All parts remaining after sampling were separated into leaf. stem and root fractions and used for fresh weight biomass determinations. Dry weight was determined after drying the samples at 55°C for 5 days. Dormant plants consisting

of three controls and five transgenes (representing lines 2, 11 and 14) were sampled at internodes 1, 10 and 20, counted from the reference point. These samples were used for maceration and subsequent fibre length measurements.

Anatomical Characterisation

Following conventional chemical fixation samples were embedded in LR White [Regan, S., Bourquin, V., Tuominen, H. & Sundberg, B. Accurate and high resolution in situ hybridization analysis of gene expression in secondary stem tissues. *Plant J.* 19, 363-369 (1999)]. Longitudinal sections for cell counts (internode 15 and 16) and transverse sections for fibre count (internode 30), 2 µm thick, were obtained and stained with toluidine blue. Numbers of epidermal and pith cells in 1 mm of the longitudinal sections were determined for 3 sections per internode. Cell length and number of cells per internode were calculated as the averages found in internodes 15 and 16. The number of xylem fibres was counted in three separate radial files (from pith to cambium) per individual and the average for each genotype (line 4, 11, 13 and control) was calculated.

Fibre Length Measurements

For fibre length determination, trimmed pieces of outer xylem from the selected internodes were prepared. The samples were macerated by boiling in a solution of 10% hydrogen peroxide and 50% glacial acetic acid for 4-6 h. When totally bleached, they were rinsed with distilled water three times, neutralised with sodium carbonate and washed again in water. Finally, the fibres were separated from each other by shaking in water and measured using a fibre analyser (KAJAANI FiberLab, Valmet Automation Kajaani Ltd, Kajaani, Finland). On average, the lengths of 30 800 fibres per sample were measured.

Northern Analysis

Total RNA was extracted from leaves using a chloroform and hexadecyltrimethylamonium bromide based method according to Chang et al. [Chang, S., Puryear, J. & Cairney, J. A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* 11, 113-116 (1993)]. About 18 µg of total RNA per sample

was separated in a formaldehyde agarose gel according to Sambrook *et al.* [Sambrook, J., Fritsch, E. & Maniatis, T. *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 1989], and blotted on to a nylon Hybond-N membrane (Amersham, Little Chalfont, UK) according to the manufacturer's protocol. A *Bam HI/Xho I* fragment recovered from pOK12:AtGA20ox1 and an ubiquitin-like Expressed Sequence Tag (EST A046p57) obtained in the *Populus* sequencing project [Sterky, F. *et al.* Gene discovery in the wood-forming tissues of poplar: Analysis of 5.692 expressed sequence tags. *Proc. Natl. Acad. Sci. USA* **95**, 13330-13335 (1998)] were used as probes for analysis of ectopic *AtGA20ox1* and reference gene expression, respectively. Hybridisation was performed in Church buffer [Church, G.M. & Gilbert, W., Genomic sequencing, *Proc. Natl. Acad. Sci. USA* **81**, 1991-1995 (1984)] at 65°C overnight. The membrane was washed at 65°C, two times three minutes in each of solution: 2 x SSPE, 0.1% SDS; 1 x SSPE, 0.1% SDS; 0.5 x SSPE, 0.1% SDS and 0.1 x SSPE, 0.1% SDS. To visualise the expression patterns, the membrane was exposed to a phosphor imager (Molecular Imager GS-525, Bio-Rad Laboratories, Hercules, CA, USA) overnight.

Quantification of Gibberellins

Samples of 200-300 mg were ground in liquid nitrogen to a homogenous powder, and GAs were analysed as earlier described by Peng *et al.* [Peng, J.R., Richards, D.E., Moritz, T., CanoDelgado, A. & Harberd, N.P., Extragenic suppressors of the Arabidopsis gai mutation alter the dose-response relationship of diverse gibberellin responses. *Plant Physiol.* 119, 1199-1207 (1999)] with a few modifications. Volumes of the solvents used for extraction and partitioning, and that of the anion-exchange column, were reduced. Furthermore, the samples were methylated prior to the HPLC step, and finally they were analysed by GC/MS-selected reaction monitoring (SRM) using a JEOL SX/SX102A four sector mass spectrometer (JEOL, Tokyo, Japan). [$^2H_2$]-GAs were used as internal standards.

Results

Generation of transgenic GA 20-oxidase hybrid aspen

Out of 14 independently transformed lines, 10 were regenerated and selected for further analysis. Southern analysis showed that all the selected transgenic lines have one to three copies of the *AtGA20ox1* gene inserted into their genomes, except for line 6, which has at least 4 copies (data not shown). Northern analysis of RNA isolated from leaf tissue showed that the strongest expression of the *AtGA20ox1* gene was in lines 2, 7, 9, 14 and 15. Expression was slightly lower in lines 1, 4, 11 and 13, while expression was almost undetectable in line 6 (Fig. 2). In general, the transgenic lines showed poorer rooting than the control plants, both under tissue culture conditions and when potted in soil. When potted, for example, 100% of the control plants survived, compared to only 32% of the transgenic plants.

GA levels

GA content was determined both in leaves and internodes of actively growing tissue. Transgenic plants showed high levels of the 13-hydroxylated C_{19}-GAs (GA_{20}, GA_1 and GA_8) and the non 13-hydroxylated C_{19}-GAs (GA_9, GA_4 and GA_{34}) in both stem and leaves (Table 1). The increase over wild type levels was more pronounced in stem tissue than in leaf tissue. The levels of the biologically active GA_1 and GA_4 in stem tissues of lines 7 were 22- and 24- fold higher, respectively, than in the control. Furthermore, all transgenic lines showed lower levels of the substrates for the GA 20-oxidase. For example GA_{12} and GA_{53} levels in stem tissues of line 7 were 17 % and 9 % of the contents in the control, respectively.

Shoot growth and morphology

All transgenic lines showed faster than wild type height and diameter growth (Fig. 3), giving the trees a characteristic phenotype (Fig. 4). Although levels of GA 20-oxidase mRNA varied between the transgenic lines (Fig. 2) there were no strict correlation between expression levels and growth. A detailed study of two of the transgenic lines (Table 2) revealed that the difference in height growth was primarily due to differences in internode lengths. No statistically significant differences in the

number of nodes (leaves) were observed (data not shown). However, leaf development was different in the transgenic lines. Young expanding leaves had a different morphology and a higher leaf length to width ratio than leaves of control plants (Table 2). After node 14, i.e. in the fully expanded leaves, there was no significant difference in this morphological ratio, but there was still a clear difference in leaf size between control and transgenic plants. The transgenics had longer and broader leaves and, consequently, higher mean leaf fresh weights (Table 2). Furthermore, the petioles were longer in the transgenic lines than in the control (data not shown).

There was a significant difference in shoot biomass between transgenic and control plants when measured either on a fresh or dry weight basis (Table 2). The transgenic plants had 64% higher shoot dry weight, on average, than the control plants. This difference was especially pronounced in the stem, where the transgenic plants had 126 % higher dry weight than the control plants. Consequently, the transgenic plants also had an altered root:shoot weight ratio, of only about 2:1 compared to 4:1 for control plants. Although the initial rooting capacity was lower in the transgenic lines, there was no significant difference in root dry weight between transgenic and control plants (Table 2).

Cell length and cell numbers in transgenic plants

Expression of the *AtGA20ox1* gene did not effect the lengths of epidermal and pith cells in presumably fully elongated internodes (see Experimental Protocol for definition) in the actively growing transgenic lines (Figure 5A). However, epidermal and pith cell numbers in these elongated internodes were approximately 55 % higher than in the wild type control (Figure 5B).

Cambial activity was also determined by counting numbers of xylem fibre cells in transverse sections (Fig. 5C). The transgenic lines show 71 % increase in number of xylem fibre cells compared to the control plants. Xylem fibres were sampled at three different positions in dormant plants and their lengths were measured. No differences in fibre length from different heights in the trees were detected (data not shown), but the xylem fibres were approximately 8 % longer in the transgenic lines than in the control plants (Fig. 5C).

Although the invention has been described with regard to its preferred embodiments, which constitute the best mode presently known to the inventors, it should be understood that various changes and modifications as would be obvious to one having the ordinary skill in this art may be made without departing from the scope of the invention as set forth in the claims appended hereto.

SEQUENCE LISTING

<110> Swetree Genomics AB

<120> Transgenic trees and methods for their production

<130> MH42953

<140>
<141>

<160> 2

<170> PatentIn Ver. 2.1

<210> 1
<211> 1259
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arabidopsis GA
20-oxidase cDNA construct

<400> 1
aatctctcaa aatggccgta agtttcgtaa caacatctcc tgaggaagaa gacaaaccga 60
agctaggcct tggaaatatt caaactccgt taatcttcaa cccttcaatg cttaaccttc 120
aagccaatat cccaaaccaa ttcatctggc ctgacgacga aaaaccttcc atcaacgttc 180
tcgagcttga tgttcctctc atcgaccttc aaaaccttct ctctgatcca tcctccactt 240
tagatgcttc gagactgatc tctgaggcct gtaagaagca cggtttcttc ctcgtggtca 300
atcacggcat cagcgaggag cttatatcag acgctcatga atacacgagc cgcttctttg 360
atatgcctct ctccgaaaaa cagagggttc ttagaaaatc cggtgagagt gttggctacg 420
caagcagttt caccggacgc ttctccacca agcttccatg gaaggagacc ctttctttcc 480
ggttttgcga cgacatgagc cgctcaaaat ccgttcaaga ttacttctgc gatgcgttgg 540
gacatgggtt tcagccattt gggaaggtgt atcaagagta ttgtgaagca atgagttctc 600
tatcattgaa gatcatggag cttctcgggc taagtttagg cgtaaaacgg gactactttа 660
gagagttttt cgaagaaaac gattcaataa tgagactgaa ttactaccct ccatgtataa 720
aaccagatct cacactagga acaggacctc attgtgatcc aacatctctt accatccttc 780
accaagacca tgttaatggc cttcaagtct ttgtggaaaa tcaatggcgc tccattcgtc 840
ccaaccccaa ggcctttgtg gtcaatatcg gcgatacttt catggctcta tcgaacgata 900
gatacaagag ctgcttgcac cgggcggtgg tgaacagcaa gagcgagagg aagtcacttg 960
cattcttctt gtgtccgaaa aaagacagag tagtgacgcc accgagagag cttttggaca 1020
gcatcacatc aagaagatac cctgacttca catggtctat gttccttgag ttcactcaga 1080
aacattatag agcagacatg aacactctcc aagccttttc agattggctc accaaaccca 1140
tctaagaaat aaaatattca tgtcttgtct tgttagttac tagtatcttc tttatatttc 1200
atgtatgtat atggtaatag gcaataacac cttttagcat ctcaaaaaaa aaaaaaaaa 1259

Leu Gly Val Lys Arg Asp Tyr Phe Arg Glu Phe Phe Glu Glu Asn Asp
210 215 220

Ser Ile Met Arg Leu Asn Tyr Tyr Pro Pro Cys Ile Lys Pro Asp Leu
225 230 235 240

Thr Leu Gly Thr Gly Pro His Cys Asp Pro Thr Ser Leu Thr Ile Leu
245 250 255

His Gln Asp His Val Asn Gly Leu Gln Val Phe Val Glu Asn Gln Trp
260 265 270

Arg Ser Ile Arg Pro Asn Pro Lys Ala Phe Val Val Asn Ile Gly Asp
275 280 285

Thr Phe Met Ala Leu Ser Asn Asp Arg Tyr Lys Ser Cys Leu His Arg
290 295 300

Ala Val Val Asn Ser Lys Ser Glu Arg Lys Ser Leu Ala Phe Phe Leu
305 310 315 320

Cys Pro Lys Lys Asp Arg Val Val Thr Pro Pro Arg Glu Leu Leu Asp
325 330 335

Ser Ile Thr Ser Arg Arg Tyr Pro Asp Phe Thr Trp Ser Met Phe Leu
340 345 350

Glu Phe Thr Gln Lys His Tyr Arg Ala Asp Met Asn Thr Leu Gln Ala
355 360 365

Phe Ser Asp Trp Leu Thr Lys Pro Ile
370 375

<210> 2
<211> 377
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arabidopsis GA 20-oxidase cDNA construct

<400> 2

```
Met Ala Val Ser Phe Val Thr Thr Ser Pro Glu Glu Glu Asp Lys Pro
 1               5                  10                  15

Lys Leu Gly Leu Gly Asn Ile Gln Thr Pro Leu Ile Phe Asn Pro Ser
            20                  25                  30

Met Leu Asn Leu Gln Ala Asn Ile Pro Asn Gln Phe Ile Trp Pro Asp
        35                  40                  45

Asp Glu Lys Pro Ser Ile Asn Val Leu Glu Leu Asp Val Pro Leu Ile
    50                  55                  60

Asp Leu Gln Asn Leu Leu Ser Asp Pro Ser Ser Thr Leu Asp Ala Ser
65                  70                  75                  80

Arg Leu Ile Ser Glu Ala Cys Lys Lys His Gly Phe Phe Leu Val Val
                85                  90                  95

Asn His Gly Ile Ser Glu Glu Leu Ile Ser Asp Ala His Glu Tyr Thr
            100                 105                 110

Ser Arg Phe Phe Asp Met Pro Leu Ser Glu Lys Gln Arg Val Leu Arg
        115                 120                 125

Lys Ser Gly Glu Ser Val Gly Tyr Ala Ser Ser Phe Thr Gly Arg Phe
    130                 135                 140

Ser Thr Lys Leu Pro Trp Lys Glu Thr Leu Ser Phe Arg Phe Cys Asp
145                 150                 155                 160

Asp Met Ser Arg Ser Lys Ser Val Gln Asp Tyr Phe Cys Asp Ala Leu
                165                 170                 175

Gly His Gly Phe Gln Pro Phe Gly Lys Val Tyr Gln Glu Tyr Cys Glu
            180                 185                 190

Ala Met Ser Ser Leu Ser Leu Lys Ile Met Glu Leu Leu Gly Leu Ser
        195                 200                 205
```

Claims

1. A transgenic woody plant exhibiting at least one improved growth parameter, **characterized** in that a DNA sequence coding for the expression of a polypeptide exhibiting GA 20-oxidase activity is functionally inserted in the plant genome.

2. The transgenic woody plant according to claim 1, **characterized** in that the improved growth parameter is one of increased growth rate, biomass production and/or xylem fibre length.

3. A transgenic woody plant of the *Populus* species exhibiting at least one improved growth parameter, **characterized** in that a DNA sequence coding for the expression of GA 20-oxidase is functionally inserted in the plant genome.

4. The transgenic *Populus sp.* plant according to claim 3, **characterized** in that the improved growth parameter is one of increased growth rate, biomass production and/or xylem fibre length.

5. A transgenic *Populus sp.* plant belonging to the species *Populus tremula x P. tremuloides* exhibiting at least one improved growth parameter, **characterized** in that a DNA sequence coding for the expression of GA 20-oxidase is functionally inserted in the plant genome.

6. The transgenic *Populus sp.* plant according to claim 5, **characterized** in that the improved growth parameter is one of increased growth rate, biomass production and/or xylem fibre length.

7. Seeds of a transgenic plant according to any one of claims 1 – 6.

8. Cells of a transgenic plant according to any one of claims 1 - 6.

9. Propagating material of a transgenic plant according to any one of claims 1 - 6.

10. A transgenic plant according to any one of claims 1 – 6. **characterized** in that the DNA sequence coding for the expression of a polypeptide exhibiting GA 20-oxidase activity is SEQ. ID. NO. 1 or a homologue thereof.

11. Seeds of a transgenic plant according to claim 7. **characterized** in that the DNA sequence coding for the expression of a polypeptide exhibiting GA 20-oxidase activity is SEQ. ID. NO. 1 or a homologue thereof.

12. Cells of a transgenic plant according to claim 8. **characterized** in that the DNA sequence coding for the expression of a polypeptide exhibiting GA 20-oxidase activity is SEQ. ID. NO. 1 or a homologue thereof.

13. Propagating material of a transgenic plant according to claim 9. **characterized** in that the DNA sequence coding for the expression of a polypeptide exhibiting GA 20-oxidase activity is SEQ. ID. NO. 1 or a homologue thereof.

14. A method of increasing at least one growth parameter in trees. **characterized** in that a DNA sequence coding for the expression of a polypeptide exhibiting GA 20-oxidase activity is functionally inserted in the plant genome.

15. The method according to claim 14. **characterized** in that the growth parameter is one of growth rate. biomass production and xylem fibre length.

16. The method according to claim 14. **characterized** in that a DNA sequence coding for the expression of a polypeptide exhibiting GA 20-oxidase activity is the sequence disclosed as SEQ. ID. NO. 1 or a homologue thereof.

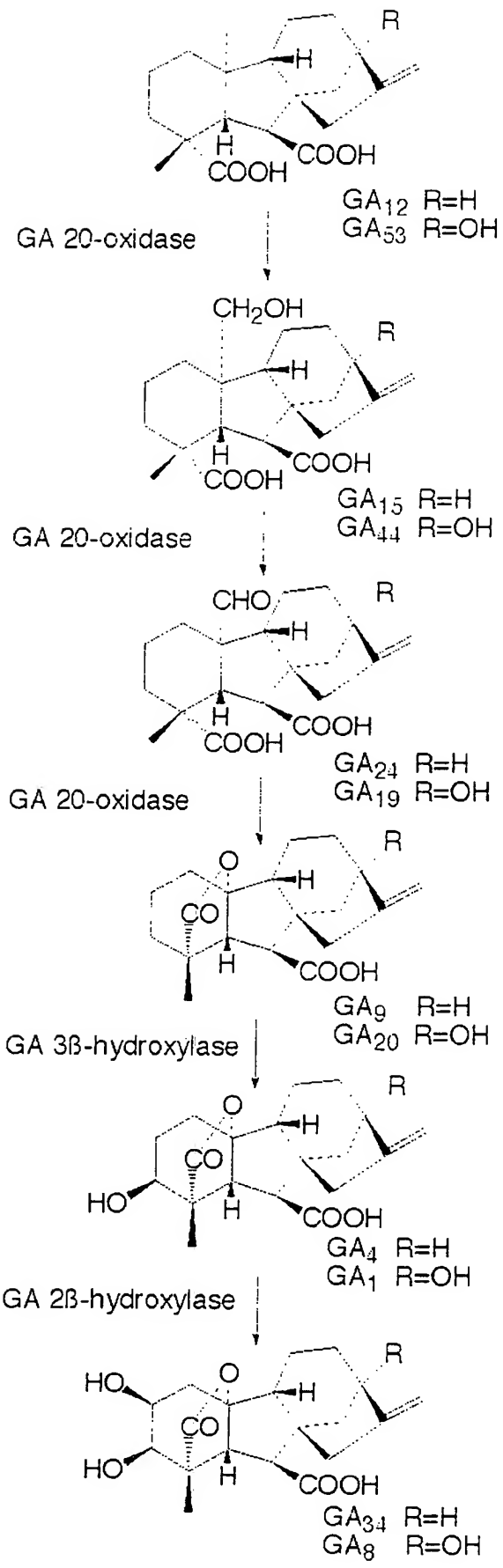

R
H
H
COOH
COOH
GA12 R=H
GA53 R=OH
GA 20-oxidase
CH2OH
R
H
H
COOH
COOH
GA15 R=H
GA44 R=OH
GA 20-oxidase
CHO
R
H
H
COOH
COOH
GA24 R=H
GA19 R=OH
GA 20-oxidase
O
R
H
CO
H
COOH
GA9 R=H
GA20 R=OH
GA 3ß-hydroxylase
O
R
H
CO
HO
H
COOH
GA4 R=H
GA1 R=OH
GA 2ß-hydroxylase
HO
O
R
H
CO
HO
H
COOH
GA34 R=H
GA8 R=OH

Fig. 1

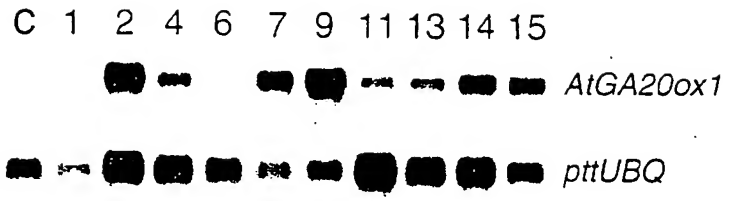


Fig. 2

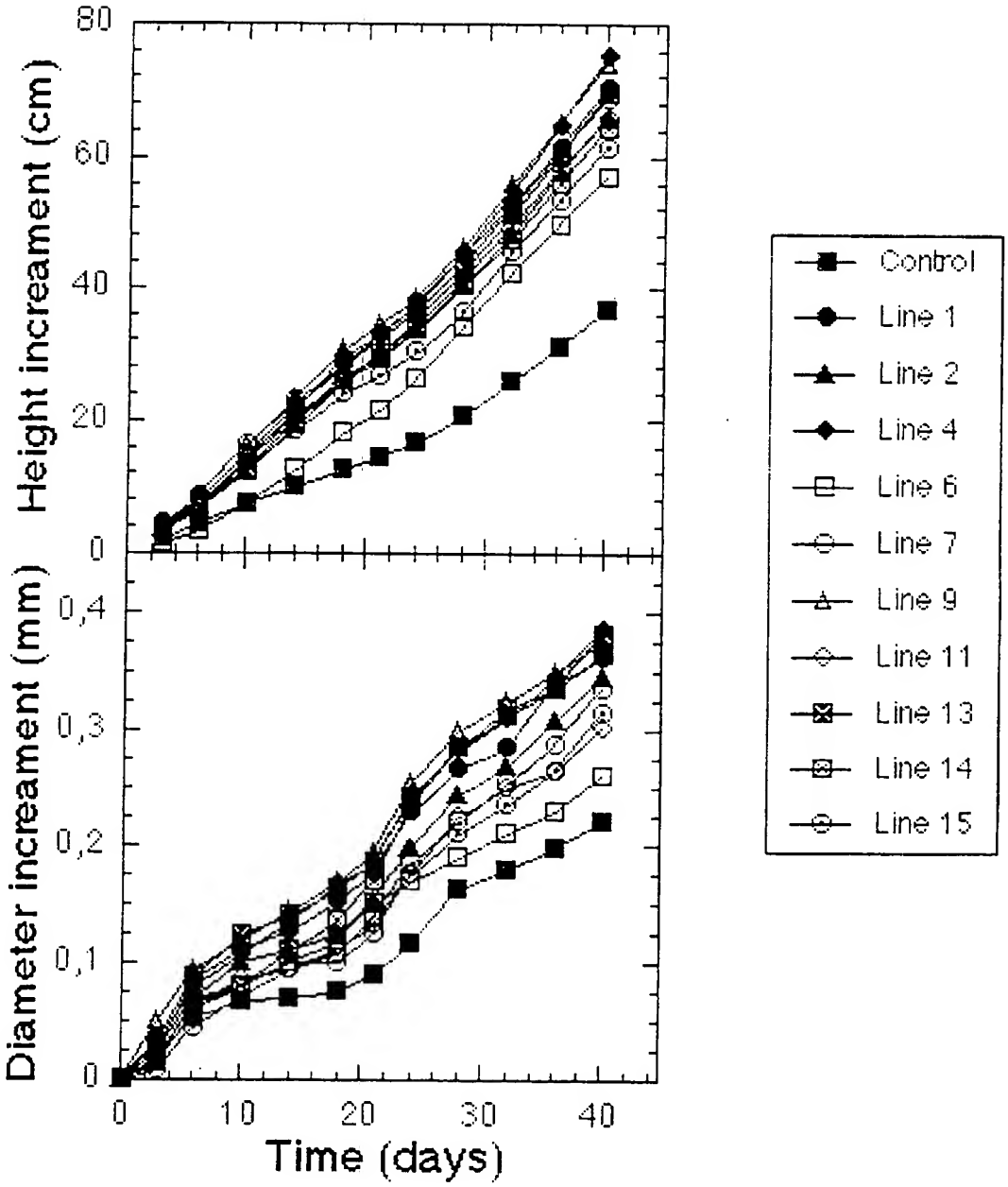

Height increament (cm)
Diameter increament (mm)
Time (days)
Control
Line 1
Line 2
Line 4
Line 6
Line 7
Line 9
Line 11
Line 13
Line 14
Line 15

Fig. 3

Fig. 4

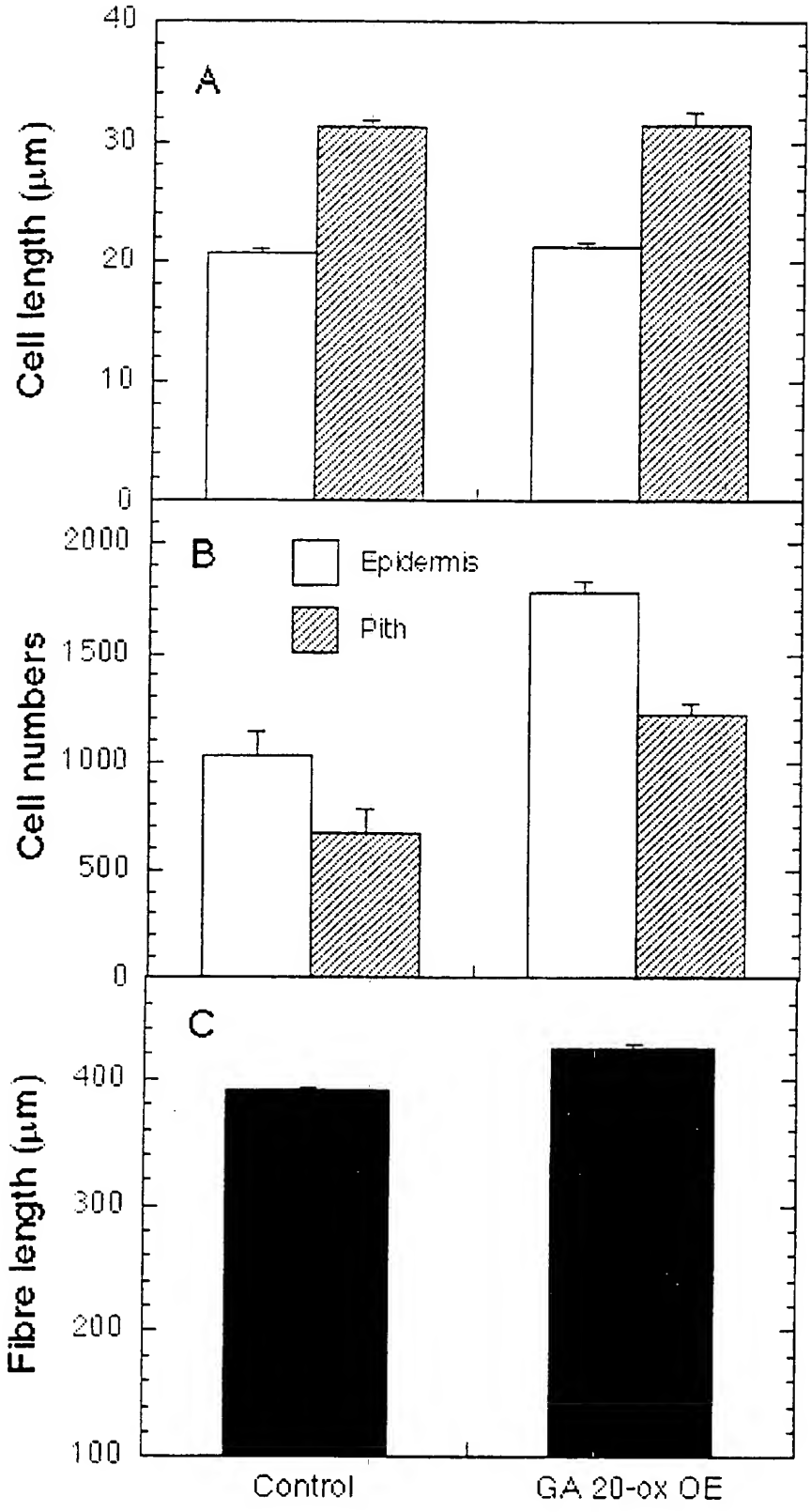

A
Cell length (µm)
40
30
20
10
0
B
Epidermis
Pith
Cell numbers
2000
1500
1000
500
0
C
Fibre length (µm)
400
300
200
100
Control
GA 20-ox OE

Fig. 5

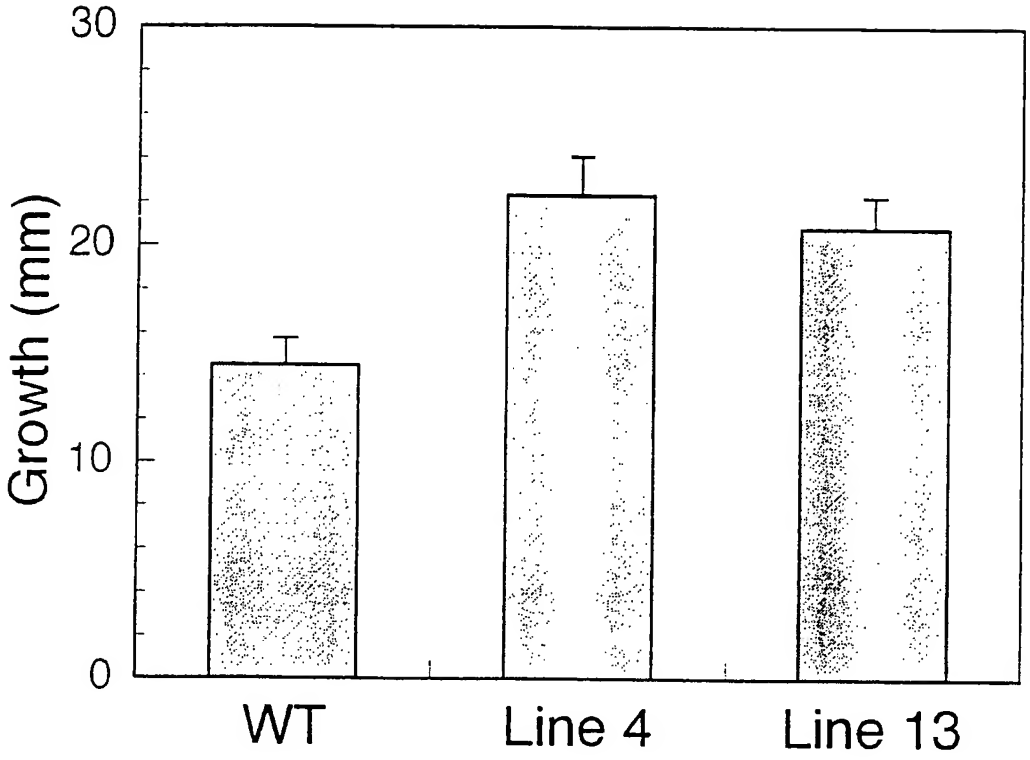


Fig. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 01/00451

A. CLASSIFICATION OF SUBJECT MATTER

IPC7: C12N 15/82, A01H 5/00
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC7: C12N, A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Plant Physiol., Volume 118, 1998, Shihshieh Huang et al, "Overexpression of 20-Oxidase Confers a Gibberellin-Overproduction Phenotype in Arabidopsis", page 773 - page 781, page 780, column 2; line 34 - 41 --	1-16
P,X	Nature America Inc., Volume 18, July 2000, Maria E. Eriksson et al, "Increased gibberellin biosynthesis in transgenic trees promotes growth, biomass production and xylem fiber length" page 784 - page 788 --	1-16
A	WO 9428141 A1 (LONG ASHTON RESEARCH STATION), 8 December 1994 (08.12.94), page 1626, line 27 - line 28, claim 35 --	1-16

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 July 2001	27 -07- 2001
Name and mailing address of the ISA Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Patrick Andersson/BS Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORT

Information on patent family members

02/07/01

International application No.
PCT/SE 01/00451

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9428141 A1	08/12/94	AU 6929794 A	20/12/94
		CA 2163454 A	08/12/94
		EP 0703983 A	03/04/96
		GB 9311147 D	00/00/00
		JP 8510381 T	05/11/96
		US 5939539 A	17/08/99
		US 6198021 B	06/03/01